

Effectiveness of Automatic Translations for Cross-Lingual Ontology Mapping

Mamoun Abu Helou

*Department of Informatics,
Systems and Communication
University of Milan-Bicocca*

MAMOUN.ABUHELOU@DISCO.UNIMIB.IT

Matteo Palmonari

*Department of Informatics,
Systems and Communication
University of Milan-Bicocca*

MATTEO.PALMONARI@DISCO.UNIMIB.IT

Mustafa Jarrar

*Department of Computer Science
Birzeit University*

MJARRAR@BIRZEIT.EDU

Abstract

Accessing or integrating data lexicalized in different languages is a challenge. Multilingual lexical resources play a fundamental role in reducing the language barriers to map concepts lexicalized in different languages. In this paper we present a large-scale study on the effectiveness of automatic translations to support two key cross-lingual ontology mapping tasks: the retrieval of candidate matches and the selection of the correct matches for inclusion in the final alignment. We conduct our experiments using four different large gold standards, each one consisting of a pair of mapped wordnets, to cover four different families of languages. We categorize concepts based on their lexicalization (type of words, synonym richness, position in a subconcept graph) and analyze their distributions in the gold standards. Leveraging this categorization, we measure several aspects of translation effectiveness, such as word-translation correctness, word sense coverage, synset and synonym coverage. Finally, we thoroughly discuss several findings of our study, which we believe are helpful for the design of more sophisticated cross-lingual mapping algorithms.

1. Introduction

Different ontology representation models have been proposed to ease data exchange and integration across applications. *Axiomatic ontologies* are represented in logic-based languages, e.g., OWL (2004), and define concepts by means of logical axioms. *Lexical ontologies* define the meaning of concepts by taking into account the words that are used to express them (Hirst, 2004): each concept is defined by one or more synonym *words* (Miller, 1995), which we refer to as *lexicalization of the concept*, and connected to other concepts by semantic relations. Several hybridizations of these two approaches have also been proposed (Vossen et al., 2010).

When data sources using different ontologies have to be integrated, mappings between the concepts described in these ontologies have to be established. This task is also called *ontology mapping*. Automatic ontology mapping methods are introduced to ease this task by finding potential mappings and determining which ones should be included in a final alignment. Ontology mapping methods perform two main sub tasks: in *candidate match*

retrieval, a first set of potential matches is found; in *mapping selection*, a subset of the potential matches is included in a final alignment.

Some aspects of concept modelling are common to lexical and logical ontologies, despite their differences: concepts have a lexicalization and are organized in subconcept graphs. *Synonymful lexicalizations*, i.e., lexicalizations that contain more than one synonym words, are more frequently found in lexical ontologies than in axiomatic ontologies. However, enriching the lexicalization of concepts in axiomatic ontologies with a set of synonyms is a well-established practice in ontology mapping (Sorrentino, Bergamaschi, Gawinecki, & Po, 2010; Shvaiko & Euzenat, 2013; Faria, Martins, Nanavaty, Taheri, Pesquita, Santos, Cruz, & Couto, 2014).

Cross-lingual ontology mapping is the task of establishing mappings between concepts of a source ontology lexicalized in a language and concepts of a target ontology lexicalized in a different language (Spohr, Hollink, & Cimiano, 2011). If we consider that more than a million datasets have been published online as linked open data in 24 different languages (LOGD, 2015), cross-lingual ontology mapping is currently considered an important challenge (Gracia, Montiel-Ponsoda, Cimiano, Gómez-Pérez, Buitelaar, & McCrae, 2012). For instance, in the COMSODE project (2015), several tables lexicalized in different languages have been published in RDF (2014) after being annotated using domain ontologies. Data publishers would like to annotate their data using ontologies lexicalized in their native language as well as in English. Annotations in the native language of the publishers can facilitate the access to local citizens, while annotations in English support the integration of data published in different countries with the large amount of data published in English. A cross-lingual ontology mapping system may help them by facilitating bilingual data annotation.

Cross-lingual ontology mapping methods are also helpful in the construction of multilingual or large lexical ontologies (De Melo & Weikum, 2009; Abu Helou, Palmonari, Jarrar, & Fellbaum, 2014). For example, in the Arabic Ontology project (Birzeit, 2011; Jarrar, 2011, 2006; Jarrar et al., 2014), a kernel of core concepts could be extended by mapping new concepts defined by synsets and glosses to the English WordNet (Miller, 1995; Fellbaum, 1998) so as to derive novel semantic relations.

Most of the cross-lingual ontology mapping methods include a step in which the concepts' lexicalizations of one ontology are automatically translated into the language of the other ontology (Pianta, Bentivogli, & Girardi, 2002; Vossen, 2004). The most frequently adopted approach to obtain automatic translations is to use multilingual *lexical resources*, such as machine translation tools or bilingual dictionaries. The quality of the translations used by a mapping method has a major impact on its performance. However, we found that a systematic and large-scale analysis of the effectiveness of automatic translations in the context of cross-lingual mapping is missing. The study presented in this paper aims at providing a significant contribution to fill in this gap.

In our study, we use two multilingual lexical resources as sources of translations: *Google Translate* (2015) and *BabelNet*¹ (Navigli & Ponzetto, 2012). Google Translate is a machine translation tool that has been frequently used in cross-lingual ontology mapping (Shvaiko, Euzenat, Mao, Jiménez-Ruiz, Li, & Ngonga, 2014). Previous work has suggested that

1. We used BabelNet version 2.5. Recent versions were released while writing this paper (BabelNet, 2012).

Google Translate performs better than other Web translation services in the context of concept mapping (Al-Kabi, Hailat, Al-Shawakfa, & Alsmadi, 2013; Oliver & Climent, 2012). In addition, the service can be configured so as to obtain *reverse translations*, which further increase the number of words that can be automatically translated. BabelNet is the largest multilingual knowledge resource available as of today. Its concepts are derived from the fusion of the English WordNet - the largest lexical ontology - and a large source of encyclopedic knowledge such as Wikipedia (2015b). Multilingual lexicalizations have been created using inter-lingual links in Wikipedia, different translation strategies and several bilingual dictionaries collaboratively created on the Web, as we will explain in detail in Sections 3.3 and 5. On the one hand, we expect that translations obtained from BabelNet cover a large number of words. On the other hand, by evaluating translations obtained from BabelNet we are indirectly evaluating different sources of translations, some of which have been used individually in several cross-lingual mapping approaches. Another reason for choosing these resources in our study is the very large number of languages covered by Google Translate and BabelNet (respectively, 90 and 272), if compared to other resources of the same kind.

Our study is organized as follows. By focusing on concepts' lexicalizations, we consider concepts as *synsets*, i.e., sets of words with an equivalent meaning in a given context (Miller, 1995). This definition is used to classify concepts (synsets) into different categories, based on different characteristics: word ambiguity (e.g., monosemous vs. polysemous), number of synonyms (e.g., synonymful vs. synonymless), and position in a concept hierarchy (e.g., leaves vs. intermediate concepts). Using these classifications, we evaluate the effectiveness of translations obtained with multilingual lexical resources by studying the performance on the cross-lingual mapping tasks executed using automatic translations for different categories of synsets. We first analyze the coverage of translations and its impact on the candidate match retrieval task. Then we analyze the difficulty of the mapping selection task using a baseline mapping selection method. These analyses are based on different measures introduced to evaluate the translation effectiveness in terms of coverage and correctness, which are based on a comparison with translations considered *perfect* according to a gold standard, i.e., a set of cross-lingual mappings that are deemed to be correct.

As gold standards, we use cross-lingual mappings manually established (or validated) by lexicographers between four wordnets (Arabic, Italian, Slovene and Spanish) and the English WordNet. Using gold standards based on these wordnets has two main advantages. They contain a large number of mapped concepts, much larger, e.g., than the gold standards used to evaluate cross-lingual ontology mapping systems in the Ontology Alignment Evaluation Initiative (OAEI) (Shvaiko et al., 2014; OAEI, 2015), and we can leverage our lexical characterization of concepts into different categories to provide a more in-depth analysis. The wordnets used in our experiments are also representative of different families of languages and of different ontology sizes.

To the best of our knowledge, this is the first attempt to carry out a systematic and large-scale study on the effectiveness of multilingual lexical resources as sources of translations in the context of cross-lingual ontology mapping. In previous work, these resources have been mostly evaluated in the context of specific algorithms (Fu, Brennan, & O'Sullivan, 2012; Spohr et al., 2011), with a limited number of gold standards, and for a limited number of languages. Our experiments lead to interesting findings, which are discussed in a (numbered) list of observations and summarized in a "lessons learned" section. Overall,

we believe that the findings of our study can be useful for the definition of more accurate and flexible mapping algorithms, based on the characterization of concepts’ lexicalization.

The paper is structured as follows. In Section 2, we introduce some preliminary definitions used in the rest of the paper. In Section 3, we overview related work, with the goal of discussing the role of automatic translations in cross-lingual ontology mapping and related fields. The evaluation measures and the multilingual lexical resources used in our study to obtain translations are presented respectively in sections 4 and 5. In section 6, we present the experiments. Conclusions and future work end the paper.

2. Preliminaries

In this section we introduce the definitions used in our study, which cover concept lexicalizations, cross-lingual mapping and translation tasks.

2.1 Lexicalization of Concepts

We consider a general definition of ontologies, focusing on the lexical characterization of concepts, and on the relations between natural language words used in concepts. This choice is motivated by the observation that even ontology matching systems that look into the semantics of axiomatic ontologies, e.g., LogMap (Jiménez-Ruiz & Grau, 2011), use concept lexicalizations to retrieve candidate matches for concepts in a source ontology. For this reason, we borrow several definitions from lexical ontologies like WordNet (Miller, 1995) and use their terminology throughout the paper.

Slightly abusing the terminology (but coherently with WordNet), **words** are lexemes associated with a concept. A word is called *simple* when it contains one token, e.g., “table”, and is called *collection*² when it contains more tokens, e.g., “tabular array”.

Wordnet organizes natural language words into *synonym* sets, so-called **synsets**. Each synset represents one underlying *concept*, i.e., a set of words (synonyms) that share the same meaning in a given context. If W is the set of words represented in a wordnet, a synset $s \subseteq \mathcal{P}(W)$ is a set of words $s = \{w_1, \dots, w_n\}$.

A synset can contain one word (*synonymless*) or many words (*synonymful*). We use “concept” and “synset” interchangeably in the rest of the paper. Depending on the specific case, we use two notations for concepts: the set notation $\{w_1, \dots, w_n\}$ is used when we need to make explicit reference to the words contained in the synset, while a symbol notation s is used when this reference is not needed. We also use the set notation $w \in s$ to state that word w is contained in synset s . The set of words contained in the concept is also called its *lexicalization*. We use a superscript to specify the natural language used in concept lexicalizations when needed, i.e., w^L , s^L , or WN^L represent a word, a synset and a wordnet respectively lexicalized in the language L .

In addition to *lexical relations*, which link individual words (e.g., *synonymy*, *antonymy*), most of wordnets support *semantic relations*, which link concepts. Hypernymy and hyponymy are the most important semantic relations in wordnets. They are defined one as

2. An alternative name used instead of *collection* is *multiword expression* (MWE), which is frequently used in particular in the literature about machine translation tool evaluation (Sag, Baldwin, Bond, Copestake, & Flickinger, 2002); we use *collection* to be coherent with WordNet terminology that is used throughout the paper.

the inverse of the other one and determine the subconcept graph in wordnets (see Section 3.1). For example, the synset {table, tabular array} is hyponym of the synset {array}, while the synset {array} is hypernym of the synset {table, tabular array}.

A word is *polysemous*, i.e., has many meanings (or *senses*) when it is a member of many synsets. In the paper the superscript “+” on the right-hand corner of a word, e.g., “*board*+”, indicates a polysemous word. A word is *monosemous*, i.e. has only one meaning when it is a member of only one synset. For example, the English WordNet³ has eight senses for the word “*table*+”; one of these senses means “a set of data arranged in rows and columns”, which has the word “*tabular array*” as a synonym word. Another sense means “food or meals in general”, which has the word “*board*+” as a synonym word.

Given the set of words W and the set of synsets S defined in a wordnet WN , the function $senses : W \mapsto \mathcal{P}(S)$ returns the set of synsets that a word belong to, defined by $senses(w) = \{s | w \in s\}$. We can now define the set of **word senses** in a wordnet, $WS = \{ \langle w, s \rangle | s \in senses(w) \}$, i.e., the set of couples $\langle w, s \rangle$, such that s is a sense of w (in a given context). Observe that the number of word senses is higher than the number of synsets because we consider all the associations between words and synsets.

Example 1. The word “*table*+” has eight senses in the English WordNet, $sense^{En}(table) = \{ \{ table^+, tabular\ array \}, \{ table^+ \}, \{ table^+ \}, \{ mesa^+, table^+ \}, \{ table^+ \}, \{ board^+, table^+ \}, \{ postpone, prorogue^+, hold\ over^+, put\ over^+, table^+, shelve^+, set\ back^+, defer^+, remit^+, put\ off^+ \}, \{ table^+, tabularize, tabularise, tabulate^+ \} \}$ ⁴.

2.2 Cross-Lingual Mapping

In the ontology matching field, *cross-lingual mapping* is defined as the task of finding and establishing mappings between concepts of a *source ontology* lexicalized in a language L_1 and concepts of a *target ontology* lexicalized in a language L_2 (Spohr et al., 2011). Mappings can represent different relations between source and target concepts. If we consider a specific mapping relation R , for a source concept s and a target concept t , the output of a mapping task is a set of couples $\langle s, t \rangle$, also called an alignment. A **cross-lingual mapping task** with a mapping relation R is composed of two main steps (or, sub tasks):

- **candidate match retrieval:** find, for each source concept lexicalized in L_1 , a set of target concepts lexicalized in L_2 . We call the concepts found in this task *candidate matches*.
- **mapping selection:** given a set of candidate matches $T = \{t_1, \dots, t_n\}$ (lexicalized in L_2) for a source concept s (lexicalized in L_1), select a set of concepts $T' \subseteq T$ such that, for each $t \in T'$, $R(s, t)$ holds. When $R(s, t)$ holds, we say that t is a *correct match* for s , and that $\langle s, t \rangle$ is a *correct mapping*.

For a more in depth analysis of the semantics of cross-lingual mappings, we refer to previous work (Abu Helou et al., 2014). A **gold standard alignment** (or, gold standard for short), denoted by gs , is an alignment between synsets (concepts) in two wordnets such that the mappings in the alignment are believed to be correct. In this paper, we consider only equivalence mappings, i.e., mappings that specify that a source and a target concepts

3. In the following we use WordNet version 3.0.

4. The senses definitions can be found online at <http://wordnetweb.princeton.edu/perl/webwn?s=table>

have equivalent meaning. It is often assumed that the cardinality of equivalence mappings is 1:1, meaning that for each source concept there is at most one correct match.

In a gold standard alignment with cardinality 1:1, a synset in the source language have at most one *equivalent synset* in a target language. We use the predicate symbol “ \leftrightarrow ” to indicate that two synsets are equivalent (express the same meaning) in a gold standard. Using synset mappings in a gold standard gs , we can define the *possible senses of a word* w^{L_1} in a target language L_2 , denoted by $senses_{gs}^{L_2}(w^{L_1})$, as the senses in L_2 that are equivalent to the senses of w^{L_1} in its native language L_1 :

$$senses_{gs}^{L_2}(w^{L_1}) = \{s^{L_2} \mid \exists s^{L_1} (w^{L_1} \in s^{L_1} \wedge s^{L_1} \leftrightarrow s^{L_2})\} \quad (1)$$

Observe that the candidate match retrieval step define an upper bound for the mapping selection step: a correct mapping can be selected only if the target of the mapping was retrieved as a candidate match. In addition, *mapping selection* is a form of *disambiguation task*: the correct meaning of a concept (the lexicalization of the concept), in the target language has to be chosen among different possible meanings. A larger number of candidate matches and little evidence for preferring one candidate over another are likely to make the selection problem more difficult.

2.3 Translation Tasks

Translating words of one language into words of another language is crucial in the context of cross-lingual concept mapping, and, in particular, in the candidate match retrieval step. For seek of clarity we consider two translation tasks: translations of single words and translations of synsets. Translations are based on external multilingual lexical resources, e.g., a machine translation tool or a dictionary built using multilingual lexical resources.

We define the **word-translation** of a word w^{L_1} into a target language L_2 with a resource D , as a function $wTrans_D^{L_2} : W^{L_1} \mapsto \mathcal{P}(W^{L_2})$ that maps a word w^{L_1} into sets of words in a target language L_2 .

We define the **synset-translation** of a synset s^{L_1} into a target language L_2 with a resource D , as a function $sTrans_D^{L_2} : S^{L_1} \mapsto \mathcal{P}(\mathcal{P}(W^{L_2}))$ that maps a synset s into sets of sets of words, each of which is the output of word-translation of some $w \in s$. The synset-translation function is defined as follows:

$$sTrans_D^{L_2}(s^{L_1}) = \{wTrans_D^{L_2}(w^{L_1}) \mid w \in s^{L_1}\} \quad (2)$$

Example 2. The synset-translation of the Italian synset $\{tavola^+, tabella\}^{It}$ into English can be given as follow: $sTrans_D^{En}(\{tavola^+, tabella\}^{It}) = \{wTrans_D^{En}(tavola^+, It), wTrans_D^{En}(tabella^{It})\} = \{\{table, board, plank, panel, diner, slab\}, \{table, list\}\}$.

Observe that in the definition of the synset-translation function we do not make the set union of the outputs of every word-translation applied to the words in a synset. Instead, using Eq.2, we can write the output of the synset-translation function as *multiset union* of the sets returned by every word-translation. For instance, in Example 2, $sTrans_D^{En}(\{tavola^+, tabella\}^{It}) = \{table^{(2)}, board^{(1)}, plank^{(1)}, panel^{(1)}, diner^{(1)}, slab^{(1)}\}$, superscript numbers between brackets indicate the frequency count of the words in the translation set. Similarly, “ $table^{(2)}$ ” means that the word “table” appears in two subsets, i.e., the word “table” is resulted as a translation of two synonym words in the source synset, which are

”tavola” and ”tabella”. In this way we can count the number of word-translations that produced one word in the target language for a given synset-translation. These counts can be helpful to use the results of the synset-translation to perform the mapping selection step. For example, these counts can be used to weigh the candidate matches with a majority voting approach, like the one used in our experiments in Section 6.3.2.

3. Automatic Translations in Cross-Lingual Mapping Tasks

In this section we review the use of automatically generated translations in *cross-lingual ontology mapping* and in other related tasks such as *enrichment of multilingual knowledge resources* and *cross-lingual word-sense disambiguation*.

The enrichment of multilingual knowledge resources is related to cross-lingual ontology mapping and to the findings of our study for several reasons. First, multilingual knowledge resources can be used as sources of translations in cross-lingual ontology mapping approaches. Second, the wordnets mapped to the English WordNet that we use as gold standards are multilingual knowledge resources, because their mappings represent inter-lingual links between concepts. Third, the two most frequently adopted approaches to enrich multilingual knowledge resources are based either on mapping concepts lexicalized in different languages or on translating the concepts’ lexicalizations. Since we evaluate the correctness and coverage of translations of ontology concepts, our findings are relevant also to approaches that intend to use these translations and ontology mapping methods to enrich multilingual knowledge resources.

Cross-lingual word sense disambiguation is another research field where translations have been used to solve a mapping problem, which is related to, but also quite different from, the mapping tasks considered in this study.

Before discussing related work in these different research areas we discuss the usage of the term “concept” in lexical and axiomatic ontologies. We conclude the section by presenting the contributions of this study to the evaluation of automatically generated translations in cross-lingual ontology mapping and related tasks.

3.1 Concepts in Lexical and Axiomatic Ontologies

Concepts are the constituents of thoughts (Margolis & Laurence, 2014). The relation between natural language and thought is much debated. For example, some maintain that concepts are independent from the language (Fodor, 1975; Pinker, 1994) while others believe that concepts require natural language to exist (Carruthers, 2002; Spelke, 2003). However, natural language plays a major role in expressing concepts in many computational knowledge representation systems proposed to support natural language processing, information retrieval and data integration tasks. *Ontologies* are among these computational knowledge representation systems. We distinguish between two different kinds of ontologies.

In *lexical ontologies*, the meaning of concepts is primarily defined in relation to the words that can be used to express them. For example, in order to represent the concept “table”, with reference to the object used to eat a meal, the set of words used to refer to this concept are specified. Lexical ontologies include domain thesauri, and wordnets, the most popular of which is the English WordNet (Miller, 1995; Fellbaum, 1998). In *axiomatic ontologies* (or, logical ontologies) the meaning of concepts is defined by axioms specified

in a logical language, e.g., First Order Logic, which are interpreted as constraints over mathematical structures and support automated reasoning (Horrocks, 2008). Examples of logical ontologies include web ontologies defined in RDFS (2014) or OWL, but an annotated database schema or a spreadsheet can also be considered an ontology based on this broad definition (Zhuge, Xing, & Shi, 2008; Po & Sorrentino, 2011; Mulwad, Finin, & Joshi, 2013; Zhang, 2014). For example, to represent the afore-mentioned concept “table”, we can define it as “a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs” in a logical language. The intended interpretation of this concept can be every such table that had ever existed in the world, or, more specifically, a list of products of type “table” described in a spreadsheet (Mulwad et al., 2013; Zhang, 2014).

Many hybrid approaches also exist. For example, efforts to assure certain logical properties of relations represented in lexical ontologies can be found in KYOTO (Vossen et al., 2010). YAGO is a logical ontology that integrates many concepts from the English WordNet (Suchanek, Kasneci, & Weikum, 2008). WordNet concepts used to annotate a database schema can be given a formal interpretation and used to support database integration (Sorrentino et al., 2010).

As a matter of fact, despite several differences, concepts modelled in lexical, axiomatic, or hybrid ontologies *share two important features*. First, concepts are organized in *subconcept graphs*, i.e., hierarchies, partially ordered sets, or lattices that define the relations between concepts based on their generality. These relations are referred to as subconcept relations in axiomatic ontologies, while different relations can be represented in lexical ontologies, e.g., hyponymy/hypernymy. Second, in every ontology concepts have lexical descriptions that may include a set of synonym words. Of course, while synonyms are first class citizens in lexical ontologies and are available for a large number of concepts, their availability is more limited in axiomatic ontologies. However, a step to enrich the concept lexicalizations of logical ontologies with synonyms extracted from dictionaries and other lexical resources is introduced in many ontology mapping approaches so as to exploit lexical matching algorithms (Shvaiko & Euzenat, 2013; Otero-Cerdeira, Rodriguez-Martinez, & Gmez-Rodriguez, 2015; Sorrentino et al., 2010; Faria et al., 2014).

3.2 Translations in Cross-Lingual Ontology Mapping

The majority of *ontology mapping* methods proposed in the literature have addressed the problem of mapping ontological resources lexicalized in the same natural language, called *mono-lingual ontology mapping*. Since mono-lingual matching systems cannot directly access semantic information when ontologies are lexicalized in different natural languages (Fu et al., 2012), techniques to reconcile ontologies lexicalized in different natural languages have been proposed (Gracia et al., 2012; Trojahn, Fu, Zamazal, & Ritze, 2014).

Translation-driven approaches have been used to overcome the natural language barriers by transforming a cross-lingual mapping problem into a mono-lingual one (Fu et al., 2012). Different multilingual lexical resources have been used to perform the translation tasks, including manual translations, machine translation tools, and bilingual dictionaries built from Web-based multilingual resources. For a rich classification and comparison of cross-lingual mapping systems we refer to the work of Trojahn et al. (2014).

Liang and Sini (2006) *manually* mapped the English thesaurus AGROVOC (2014) to a Chinese thesaurus CAT (2014). The mappings generated by such approaches are likely to be accurate and reliable. However, this can be a time and resource consuming process specially for maintaining large and complex ontologies.

Machine translation tools are widely adopted for cross-lingual ontology mapping. Spohr et al. (2011) translate the ontology labels into a pivot language (English) using the machine translation tool (Bing, 2016). Then, they define a feature vector based on a combination of string-based and structural-based similarity metrics and learn a matching function using a support vector machine. Like other approaches based on supervised machine learning algorithms, their approach has the disadvantage of requiring a significant number of training samples and well-designed features to achieve good performance. Fu et al. (2012) translate ontology labels using Google Translate, and then match these translated labels by combining different similarity measures. Their approach leverages structural information about the ontology concepts by considering their neighbours in the matching process. Other approaches have been proposed that also apply string-based, lexical and structural matching methods to ontology labels translated with machine translation tools, like Google Translate or Bing (Faria et al., 2014; Jiménez-Ruiz, Grau, Xia, Solimando, Chen, Cross, Gong, Zhang, & Chennai-Thiagarajan, 2014; Djeddi & Khadir, 2014).

Multilingual knowledge resources available on the web have been also exploited to translate concepts' labels (Hovy, Navigli, & Ponzetto, 2012). *Wiktionary* (2015) was used to generate translations to match English and French ontologies (Lin & Krizhanovsky, 2011). First, a bilingual English-French lexicon is built using *Wiktionary* and is used to translate the labels of the ontologies. Then, the monolingual ontology matching system COMS is used (Lin, Butters, Sandkuhl, & Ciravegna, 2010). COMS uses a set of string-based, lexical and structural matching techniques to find the appropriate mappings. A similar approach uses *Wikipedia* inter-lingual links to retrieve candidate matches for source concepts (Bouma, 2010; Hertling & Paulheim, 2012). However, when used alone, Wiktionary and Wikipedia inter-lingual links may have limited coverage, in particular for resource-poor languages.

In spite of these efforts, cross-lingual mapping systems still perform significantly worse than mono-lingual mapping systems according to recent results in the OAEI contest (Shvaiko et al., 2014), which suggest that cross-lingual ontology mapping is still a very challenging problem (Trojahn et al., 2014). The datasets used to evaluate cross-lingual mapping in the OAEI, i.e., the datasets in the multifarm track (Meilicke et al., 2012), consist of alignments established between axiomatic ontologies of relatively small size and specific to the domain of conference organization. Since in our study we want to investigate translations obtained with different multilingual lexical resources at a large scale and not in a specific domain, we decided to use different and larger gold standards in our experiments.

3.3 Translations in the Enrichment of Multilingual Knowledge Resources

Several multilingual wordnets (lexical ontologies) were developed by manually or automatically translating concepts of the English WordNet into new languages (Pianta et al., 2002; Vossen, 2004; Tufis, Cristea, & Stamou, 2004; Gonzalez-Agirre, Laparra, & Rigau, 2012; Tomaž & Fišer, 2006). The *expand* and *merge* models (Vossen, 2004) are the main approaches used in the development of multilingual wordnets. In the merge model, synsets of

a pre-existing resource in one language (e.g., a thesaurus, or even an unstructured lexical resource like a dictionary) are aligned to the most equivalent synset in English. In the expand model, English synsets are translated into the respective languages. The main advantage of these two approaches is to avoid the expensive manual elaboration of the *semantic hierarchy* in new languages. The English WordNet (Fellbaum, 1998) hierarchy is used as reference for all wordnets. Moreover, any ontology that is built following these approaches is also automatically mapped to the English WordNet.

In several wordnets, the English concepts were *manually* translated by human lexicographers using external lexical resources such as dictionaries, thesauri and taxonomies. This approach has been applied to build for example the Arabic wordnet (Rodríguez et al., 2008), the Italian wordnet (Pianta et al., 2002), the Spanish wordnet (Gonzalez-Agirre et al., 2012) and the core of the Slovene wordnet, all used in our experiments. However, the manual approach to construct ontologies that aim to cover natural languages’ lexicons is often an effort-intensive and time-consuming task (De Melo & Weikum, 2009). *Automatic* approaches have been therefore proposed to reduce the lexicographers’ workload.

Parallel corpora have been used in building wordnets for languages other than English. The basic assumption underlying these methods is that the translations of words in real texts offer insights into their semantics (Resnik & Yarowsky, 1999). The Slovene wordnet was enriched using word alignments generated by a sentence-aligned multilingual corpus (Fišer, 2007). The wordnet has been further extended using bilingual dictionaries and inter-lingual links in Wikipedia. A similar approach is also followed in building the French wordnet (Sagot & Fišer, 2008). The monosemous words in the English WordNet were automatically translated using bilingual French-English dictionaries built from various multilingual resources, such as Wikipedia inter-lingual links, Wiktionary, Wikispecies (2015), and the EUROVOC thesaurus (2015).

Sentence-aligned parallel corpora may not be available for all pair of all natural languages. In addition, specific tools are needed to perform sentence and/or word alignment across the corpora, and the bilingual dictionaries extracted from these corpora are biased towards the domains they cover. To overcome these limitations, in the Macedonian wordnet (Saveski & Trajkovski, 2010), a *machine translation tool* has been used to create parallel corpora. Monosemous English words were directly translated using a bilingual English-Macedonian dictionary. For polysemous words, the English WordNet sense-tagged glosses (WordNet-Princeton, 2015) was automatically translated into Macedonian using Google Translate.

A supervised method to automatically enrich English synsets with lexicalizations in other languages was also proposed (De Melo & Weikum, 2012). This method learns to determine the best translation for English synsets by taking into account bilingual dictionaries, structural information in the English WordNet, and corpus frequency information.

Other approaches to enrich multilingual knowledge resources have been proposed to build the Universal WordNet (UWN, De Melo & Weikum, 2009), WikiNet (Nastase, Strube, Boerschinger, Zirn, & Elghafari, 2010), and BabelNet (Navigli & Ponzetto, 2012), which integrate multilingual encyclopedic knowledge from Wikipedia with the English WordNet. In this paper, we focus on BabelNet, the largest multilingual knowledge resource as of today, and use it in our study to build bilingual dictionaries that we use for translation (explained

in Section 5). A comprehensive comparison amongst the afore-mentioned three resources can be found in the work of Navigli and Ponzetto (2012).

BabelNet (Navigli & Ponzetto, 2012) has been built by integrating the English WordNet with Wikipedia. These two resources have been mapped using an unsupervised approach. As a result, BabelNet covers approximately the 83% of WordNet’s nominal synsets. Synsets from the English WordNet cover in particular (but not only) classes of objects, e.g., “University”⁵, while Wikipedia entries cover in particular (but not only) named entities, e.g., “University of Milano-Bicocca”^{6,7}. Synsets from the English WordNet and other BabelNet entries are *enriched* with lexicalizations in other languages using a variety of lexical resources. A first set of lexicalizations in languages other than English are obtained by using inter-lingual links of Wikipedia. Synsets for which Wikipedia entries cannot be found have been enriched using automatic translations of English senses-tagged sentences, extracted from Wikipedia and the *SemCor* corpus (Miller, Leacock, Tengi, & Bunker, 1993). The most *frequent* translation in a given language is detected and included as a variant lexicalization in this language; this approach was named *context-translation*. Translations of *monosemous* English words have been collected using Google Translate and directly included in the expanded lexicalizations; this approach was named *contextless-translation*. Observe that contextless translations are based on an heuristics, i.e., that monosemous words are correctly translated (also referred to as monosemous word heuristics). The *core* of BabelNet consists of the lexicalizations obtained with these approaches, also named *BabelNet synsets*. Later, BabelNet synsets’ lexicalizations are expanded with more multilingual lexical resources: Wiktionary, WikiData (2015), OmegaWiki (2015), and several wordnets that are mapped to the English wordnet, which are available through the Open Multilingual Wordnet (OMWN, 2015; Bond & Foster, 2013).

BabelNet lexicalizations (synsets) have been evaluated against manually mapped wordnets, which we also use in our experiments as gold standards. They also performed a manual evaluation with a randomly sampled set of concepts. A limit of their evaluation consists in not making explicit if the sampled senses uniformly cover polysemous and monosemous senses. Otherwise this distinction is important to evaluate different translations, also because a vast number of translations have been obtained using the contextless approach, which is based on the monosemous word heuristics. In our experiments (Section 6) we specifically analyze the effectiveness of the monosemous word heuristics in the context of ontology mapping.

We observe that the expand model was used more substantially than the merge model in approaches to automate the enrichment of multilingual wordnets and knowledge resources. One may attempt to enrich an existing wordnet via the merge approach by mapping an unstructured or a weakly structured lexicon, e.g., a dictionary, to a structured reference ontology, e.g., the English WordNet. For example, in the Arabic Ontology Project (Jarar, 2011; Abu Helou et al., 2014), the authors plan to use this approach to extend a core ontology manually created and mapped to the English WordNet. However, the mapping task incorporated in this approach is particularly challenging (Abu Helou, 2014): the lack of semantic relations between the synsets of an unstructured lexicon makes it difficult to

5. <http://wordnetweb.princeton.edu/perl/webwn?s=university>

6. https://en.wikipedia.org/wiki/University_of_Milano-Bicocca

7. http://babelnet.org/synset?word=University_of_Milano-Bicocca

disambiguate their meaning during the translation and the matching steps (Shvaiko & Euzenat, 2013; Trojahn et al., 2014). An effective cross-lingual ontology mapping method can support the application of the merge model at large scale, thus supporting the construction and enrichment of multilingual knowledge resources. For example, a recent work suggests that this approach, despite the difficulty of the task, can return multilingual concept lexicalizations richer than the ones that can be obtained by automatically translating the concepts’ labels (Abu Helou & Palmonari, 2015).

3.4 Translations in Cross-Lingual Word Sense Disambiguation

Cross-lingual ontology mapping is also related to the Cross-lingual Word Sense Disambiguation problem (*CL-WSD*), which has been studied in the recent past and addressed in SemEval 2010 and 2013 challenges (Els & Véronique, 2010; Lefever & Hoste, 2013). The goal of *CL-WSD* is to predict semantically correct translations for ambiguous words in context (Resnik & Yarowsky, 1999).

In *CL-WSD*, the lexical disambiguation task is performed as a word translation task, called *lexical substitution task* (McCarthy & Navigli, 2009). Given a source word in a sentence (e.g., an Italian word), the system tries to translate the word into a different language (e.g., English). The translation is considered to be correct if it preserves the sense that the word has in its context also in the target language.

Most of *CL-WSD* systems rely on *parallel corpora* (Gale, Church, & Yarowsky, 1992; Resnik & Yarowsky, 1999; Apidianaki, 2009), including those which exploit existing multilingual wordnets (Ide, Erjavec, & Tufis, 2002). However, the success and coverage of these methods highly depends on the nature of the parallel corpora and on the way the extracted information is used to select the appropriate senses. Corpora are known to have domain-orientated coverage, i.e., fine-grained senses for different domains might not be found in specific parallel corpora (Navigli, 2009). More importantly, parallel corpora may not be available for language couples or for specific domains (Apidianaki, 2009; Saveski & Trajkovski, 2010).

One fundamental difference between the CL-WSD task and cross-lingual ontology mapping is that in CL-WSD a context is always available and defined by the sentence a word occurs in. In cross-lingual ontology mapping the context can be defined by the neighbours of a translated concept, may be limited (Mulwad et al., 2013; Zhang, 2014), or may not be even available, e.g., when an unstructured lexicon is matched against a structured ontology (Abu Helou et al., 2014).

3.5 Scope and Contribution of this Study

Even if most of the approaches to cross-lingual ontology mapping are based on transforming a cross-lingual mapping problem into a monolingual one by leveraging translations obtained from machine translations tool or multilingual lexical resources (Trojahn et al., 2014), few efforts have been dedicated to systematically study the effectiveness of these translations in cross-lingual ontology mapping.

Fu, Brennan, and O’Sullivan (2009) studied some limitations of translation-based ontology mapping approaches, in particular, to what extent inadequate translations can introduce “noise” into the subsequent mapping step or fail to cover an adequate number of

concepts. They performed two experiments, which examined mappings of independent, domain specific, and small-scale ontologies that are labeled in English and Chinese: the Semantic Web for Research Communities ontology and the ISWC ontology. The ontologies have not been lexically enriched. Fu and her colleagues classified the translation errors introduced by machine translation tools into three main categories. *Inadequate* translation, when the translation of a source concept returns a word, which belongs to a concept in the target ontology that is more specific/generic than its equivalent concept; *synonymic* translation, when the translation of a source concept returns a word, which is synonym of the word used in the target ontology to denote its equivalent concept (but different from the one that is used by the target ontology); and *incorrect* translation, when the translation is simply wrong. In addition, the study showed that translating ontology labels in isolation leads to poorly translated ontologies which then yields low-quality matching results, thus, label translations should be conducted within context. The context is characterized by the surrounding ontology concepts.

Spohr et al. (2011) observed that, when the target ontology is lexicalized in more than one language, it is convenient to translate the source concepts in each of these languages and merge the evidence provided by these translations. However, this can be applicable only when multilingual labels are available in the target ontology, which is not the case in several cross-lingual mapping scenarios. In addition, to obtain better translations the study suggested that translating the source and target ontologies' labels into a pivot language can improve, to some extent, the quality of the translation. However, the authors stated that further evidence and experiments with several language pairs are needed to support this claim, as the quality of machine translations depends significantly on the pair of considered languages.

In this paper we analyze the effectiveness of automatic translations for cross-lingual concept mapping using large scale, general domain, and lexically rich ontologies (wordnets). The ontologies used in our studies cover four different families of languages besides English. We study the effectiveness of translations by conducting a large number of experiments that address the candidate match retrieval and the mapping selection steps of an ontology mapping process. Overall, we believe that none of previous work on cross-lingual ontology mapping provided such a systematic study, if compared in terms of scale (size of the considered concepts), number of considered languages, and level of detail of the analysis (concept categorization).

The analyses discussed in this paper can be also related to the studies on automatic translation strategies conducted to evaluate BabelNet, which is one of the two multilingual lexical resources used as a source of translation in our study. In our work, we quantitatively evaluate the correctness and the coverage of the translation strategies used in BabelNet as means to support cross-lingual mapping tasks (using mappings between wordnets for comparison; see Section 6.3.1). The studies conducted to evaluate BabelNet were aimed, instead, at evaluating their translation strategies as means to enrich multilingual lexicalizations for the concepts. We introduce two new measures to evaluate the correctness and coverage of translations obtained from multilingual resources, i.e., *translation correctness* and *synonym coverage* (see Section 4.3.2). In addition, coverage and correctness of automatic translations in our study are evaluated by considering different categories of synsets (defined in

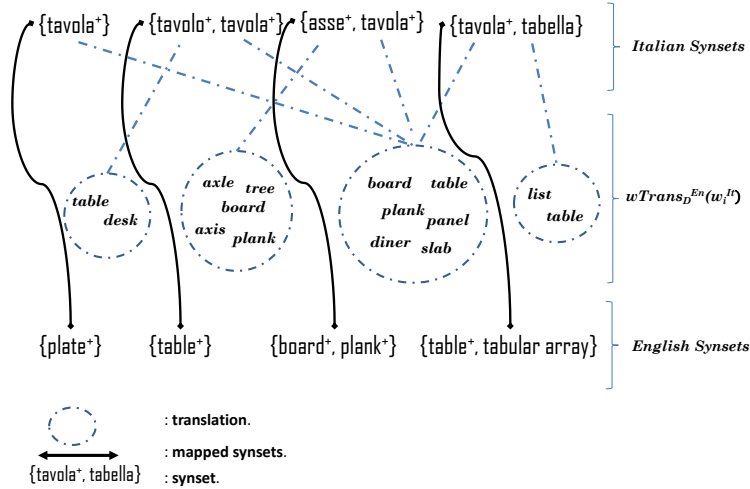


Figure 1: Example: Synset-translation

Section 6.2). Finally, we analyze the effectiveness of the monosemous word heuristic, which is used in several mapping systems and in BabelNet (contextless translation).

4. Evaluation Measures

In our study, we want to estimate the effectiveness of translations obtained from multilingual lexical resources (hereafter referred to as *resources*) in finding candidate matches for a large set of concepts. We also want to estimate the difficulty of selecting one correct mapping among a set of candidate matches, based on the information provided by translations.

For the first objective, we define four measures that we use to evaluate translation correctness and coverage. The first two measures, **translation correctness** and **word sense coverage**, are used to evaluate the effectiveness of word-translations for a given word independently of its meaning, i.e., when the sense of the word is not given. The other two measures, **synset coverage** and **synonym coverage**, are used to evaluate the effectiveness of synset-translation for a given synset focusing on the lexicalization of the synsets in the target language. Word sense coverage and synset coverage are two measures proposed in previous work by Navigli and Ponzetto (2012), but we rewrite their definitions according to the notation introduced in Section 2.3. Translation correctness and synonym coverage are introduced in this study. To facilitate the definition of these measures we first introduce the definition of *perfect translations* with respect to a gold standard. From these measures we can derive several measures, e.g., by averaging their values across one wordnet, to present the results of our experiments. For the second objective, we use a measure that is straightforwardly derived from the well-known Precision measure (Shvaiko & Euzenat, 2013) and is explained, directly, in Section 6.3.2.

4.1 Perfect Translations with a Gold Standard

The **perfect word-translation** of a word w^{L_1} into a target language L_2 w.r.t a gold standard gs is the set of every synonym words in all the possible senses of w^{L_1} in a target language L_2 :

$$wTrans_{gs}^{L_2}(w^{L_1}) = \left\{ \bigcup_{i=1}^n w_i^{L_2} \mid \exists s^{L_2} (s^{L_2} \in senses_{gs}^{L_2}(w^{L_1}) \wedge w_i^{L_1} \in s^{L_2}) \right\} \quad (3)$$

Example 3. Figure 1 illustrates the synset-translation tasks for four Italian synsets into English. Each synset is mapped to its equivalent synset in English as specified by a gold standard gs . The translations are also obtained from the mappings between the Italian and the English wordnets represented in gs . For instance, the four (Italian \leftrightarrow English) synsets mappings are: $\{tavola^+, tabella^+\} \leftrightarrow \{table^+, tabular\ array\}$, $\{asse^+, tavola^+\} \leftrightarrow \{board^+, plank^+\}$, $\{tavolo^+, tavola^+\} \leftrightarrow \{table^+\}$, and $\{tavola^+\} \leftrightarrow \{plate^+\}$. In Figure 1, the perfect word-translation of the Italian word “*tavola*” into English can be given as follow: $wTrans_{gs}^{En}(tavola^{+,It}) = \{table^+, tabular\ array, board^+, plank^+, plate^+\}^{En}$.

Observe that the perfect word-translation function returns every word of every possible sense in the target language, i.e., a word translation is perfect when it returns the complete lexicalization of every possible senses of an input word in the target language. This definition is motivated by the scope of our analysis, which evaluates the effectiveness of automatic translations in settings where the domain is not determined a-priori. When an individual input word is considered outside of a specific context, e.g., a specific sentence, a specialised domain or a concept hierarchy, the meaning of the word cannot be disambiguated, unless the word is monosemous. Otherwise, we observe that a domain-specific machine translation system, e.g., specialised in the financial domain, could determine the correct meaning (and translation) of a word, even when the word is considered individually, because of an implicit interpretation of the context by the system. Thus, in consideration of polysemous words and in absence of context specification, we defined a translation of a word (i.e., the set of words returned by a source of translation) *perfect* when it contains, for every possible usage of this word, all possible lexicalizations in the target language. If one considers word-translations in some specialized domain, he/she may need to adapt the definition of perfect word-translation consequently.

The **perfect synset-translation** of a synset s^{L_1} into a target language L_2 w.r.t a gold standard gs is defined as the set of every synonym words of the synset in L_2 mapped to s^{L_1} in gs . The perfect synset-translations can be defined as follows:

$$sTrans_{gs}^{L_2}(s^{L_1}) = \left\{ \bigcup_{i=1}^n w_i^{L_2} \mid \exists s^{L_2} (w_i^{L_2} \in s^{L_2} \wedge s^{L_1} \leftrightarrow s^{L_2}) \right\} \quad (4)$$

Example 4. In Figure 1, the perfect synset-translation of the Italian synset $\{tavola^+, tabella^+\}$ can be given as follow: $sTrans_{gs}^{En}(\{tavola^+, tabella^+\}^{It}) = \{table^+, tabular\ array\}^{En}$.

4.2 Evaluation of Word-Translations

In this section we introduce the translation correctness and the word sense coverage measures.

4.2.1 TRANSLATION CORRECTNESS

For an input word, this measure evaluates to what extent a resource returns precise and complete translations when compared to perfect word translations defined by a gold standard, which consider every possible sense of the word in the target language.

To define this measure, we need to specify when a word returned by a resource is correct. A word w^{L_2} is a **correct-translation** for a word w^{L_1} w.r.t to a gold standard gs , if w^{L_2} belongs to the set of perfect word-translations for w^{L_1} w.r.t gs (denoted by $wTrans_{gs}^{L_2}(w^{L_1})$). This principle is captured by the function $correctT_{w^{L_1},D}(w^{L_2})$ defined by the following equation:

$$correctT_{w^{L_1},D}(w^{L_2}) = \begin{cases} 1 & \text{if } w^{L_2} \in \{wTrans_{gs}^{L_2}(w^{L_1})\}. \\ 0 & \text{otherwise.} \end{cases}$$

Example 5. In Figure 1 the English words “table”, “board”, and “plank” are correct translations for the Italian word “tavola”, e.g., $correctT_{tavola^{It}}(table^{En}) = 1$. The English words “diner”, “panel”, and “slab” are *incorrect* translations for the Italian word “tavola”, e.g., $correctT_{tavola^{It}}(slab^{En}) = 0$.

We measure the correctness of translations returned by a resource D for a word w^{L_1} with *translation-correctness* as defined in Eq. 5. The measure is computed as the harmonic mean, i.e., F_1 -measure, of two measures: 1) *Precision* (Pr), defined as the number of correct translations returned by the resource D over the total number of translations returned by D ; 2) *Recall* (R)⁸, defined as the number of correct translations returned by the translation resource D over the total number of perfect word translations. We use Recall, Precision and F_1 -measure (computed with its standard range), but normalized in the range [0..100]. When no translation is returned by the resource D , Precision is set to zero.

$$Pr = \frac{|\{w^{L_2} | correctT_{w^{L_1},D}(w^{L_2})\}|}{|\{wTrans_D^{L_2}(w^{L_1})\}|} * 100, R = \frac{|\{w^{L_2} | correctT_{w^{L_1},D}(w^{L_2})\}|}{|\{wTrans_{gs}^{L_2}(w^{L_1})\}|} * 100$$

$$TransCorrectness_D^{L_2}(w^{L_1}) = F_1(Pr, R) * 100 = 2 \frac{Pr * R}{Pr + R} * 100 \quad (5)$$

Example 6. In the example shown in Figure 1, the correctness of English translation of the Italian word “tavola” is computed as follows: recall $R = 60.0$, precision $Pr = 50.0$, and the translation-correctness $TransCorrectness^{En}(tavola^{It}) = 55.0$.

4.2.2 WORD SENSE COVERAGE

For an input word, this measure evaluates how many of its possible word senses in a target language are covered at least by a word translation (as defined in Navigli & Ponzetto, 2012). A translation *covers a sense* s^{L_2} of an input word w^{L_1} in a different language when the resource returns at least one word of s^{L_2} . We use the binary predicate $cov(x, y)$ to state that a word-translation x covers the sense y . Word senses coverage tells to what extent the polysemy of a word is covered by a translation resource. Ideally, a resource is effective in

8. We remark that Recall is also named *translation accuracy* in the *WSD* literature (Navigli, 2009).

translating a word w^{L_1} when it is able to return some correct-translations for every possible sense of w^{L_1} in L_2 .

Given a word w^{L_1} translated into a target language L_2 with a resource D , the word senses coverage of w^{L_1} is defined as follows:

$$wsCoverage_D^{L_2}(w^{L_1}) = \frac{|\{s^{L_2} \mid s^{L_2} \in senses_{gs}^{L_2}(w^{L_1}) \wedge cov(wTrans_D^{L_2}(w^{L_1}), s^{L_2})\}|}{|\{s^{L_2} \mid s^{L_2} \in senses_{gs}^{L_2}(w^{L_1})\}|} \quad (6)$$

Example 7. In Figure 1, the polysemous Italian word “tavola” has four senses, each one is mapped into an equivalent synset in English. Using the translation resource D three out of the four senses are covered (Eq.6). For instance, the senses $\{table^+\}$ is covered, $cov(wTrans_D^{En}(tavola), \{table\}) = 1$, while the sense $\{plate^+\}$ is *not* covered, $cov(wTrans_D^{En}(tavola), \{plate\}) = 0$.

4.3 Evaluation of Synset-Translations

In this section we introduce the synset and synonym coverage measures.

4.3.1 SYNSET COVERAGE

This measure is defined as boolean function applied to an input synset (Navigli & Ponzetto, 2012). A synset s^{L_1} is covered if its translation, i.e., the multi set union of the translation of its constituent words, returns at least one word of its equivalent synset in the target language. This measure is useful when computed for a set of source synsets as presented by Navigli and Ponzetto. For example, by computing the percentage of source synsets mapped in a gold standard that are covered by translations obtained from a lexical resource, we can evaluate the number of mappings that can be discovered by using this translation resource.

To formally define synset coverage in a compact way, we can use the concept of perfect synset translation for a synset s^{L_1} in a target language L^{L_2} , denoted by $sTrans_D^{L_2}(s^{L_1})$. If s^{L_1} is a synset translated into a target language L_2 with a resource D , synset coverage is defined as follows:

$$sCoverage_D^{L_2}(s^{L_1}) = \begin{cases} 1 & \text{if } \exists w^{L_2} (w^{L_2} \in sTrans_D^{L_2}(s^{L_1}) \wedge w^{L_2} \in sTrans_{gs}^{L_2}(s^{L_1})) \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Example 8. Consider the Italian and their equivalent English synsets depicted in Figure 1. Three out of four Italian synsets are covered because their translation returns at least one word of their equivalent English synsets. For instance, the mapping $\{tavolo^+, tavola^+\} \leftrightarrow \{table^+\}$ is covered, while the mapping $\{tavola^+\} \leftrightarrow \{plate^+\}$ is not covered.

4.3.2 SYNONYMS COVERAGE

For an input synset s^{L_1} , this measure evaluates the number of words of s^{L_1} for which a word-translation covers the equivalent synset in the target language. This measure tells how many synonyms in a concept lexicalization are covered by correct translations.

Give a synset s^{L_1} having its equivalent synset s^{L_2} in the target language; s^{L_1} is translated using a resource D , then the synonym coverage of s^{L_1} is defined as follows:

$$\text{synonymsCoverage}_D^{L_2}(s^{L_1}) = \frac{|\{w^{L_1} \mid w^{L_1} \in s^{L_1} \wedge \text{cov}(w\text{Trans}_D^{L_2}(w^{L_1}), s^{L_2})\}|}{|s^{L_1}|} \quad (8)$$

Example 9. In Figure 1 the Italian synsets $\{\text{tavola}^+, \text{tabella}\}$, $\{\text{asse}^+, \text{tavola}^+\}$, and $\{\text{tavolo}^+, \text{tavola}^+\}$ have full synonym words coverage (Eq.8). Whereas, the synset $\{\text{tavola}^+\}$ is not covered because its only word is not covered.

Synonym coverage is a valuable measure to evaluate translations obtained from lexical resources in the field of cross-lingual concept mapping. Consider, for example, an input synset s^{L_1} and a translation resource that returns many of the (synonym) words of its equivalent synsets s^{L_2} in the target language. On the one hand, these synonym words are useful to increase the probability of finding s^{L_2} among the candidate matches of s^{L_1} . On the other hand, these synonym words can be used as evidence for selecting s^{L_2} as the best match for s^{L_1} , e.g., if compared to other candidate matches for which little evidence is collected via translation⁹. Finally, we observe that synonym words coverage is a complementary indication of the word senses coverage to measure the effectiveness of a translation resource, i.e., the coverage of the synonym words is a tool to disambiguate the polysemy of translations returned by a translation resource.

Throughout this paper, in order to quantify the overall coverage measures and correctness of the word-translation tasks across each dataset (wordnet), we compute the *Macro-average measure* (Vincent, 2013). The reported coverage measures are normalized in the range [0..100].

5. Multilingual Lexical Resources for Translation

Automatic translations can be obtained using different kinds of multilingual machine-readable lexical resources. The selection of these resources depends on the level of information they encode, for instance, the quality (accuracy) of translations they provide, the lexical domains they cover. These resources include: dictionaries, thesauri, wordnets, machine translation tools, and Web-based collaborative multilingual knowledge resources (resources in which lexical knowledge is manually and collaboratively generated, e.g., Wikipedia).

In this study two multilingual lexical resources are used as sources of translations: *Google Translate* and *BabelNet*. Google Translate is a statistical machine translation tool. Different machine translation systems exist that could be used; for instance, rule-based systems, e.g., Apertium (Apertium, 2015), and statistical-based systems, e.g., UPC (Mariò et al., 2006). We used Google Translate because previous work suggested that it performs better than other Web translation services in the context of concept mapping (Al-Kabi et al., 2013; Oliver & Climent, 2012), and has been adopted by several matching systems including the ones evaluated in the OAEI (Shvaiko et al., 2014). Moreover, Google Translate is a generic statistical machine translation, domain-independent system, and covers a very large number of languages, including the ones considered in our study. A common evaluation measure

9. This intuition has been used, for example, in a cross-lingual similarity measure proposed to support matching of lexical ontologies lexicalized in different languages (Abu Helou & Palmonari, 2015).

```

{"sentences":
  [{"trans": "table", "orig": "tavola", "translit": "", "src_translit": ""}],
  "dict":
  [{"pos": "noun", "terms": ["table", "board", "panel", "plank", "slab", "diner"],
    "entry": [
      {"word": "table", "reverse_translation": ["tabella", "tavolo", "tavola", "tavolino", "prospetto", "mensa"], "synset_id": [55801], "score": 0.443},
      {"word": "board", "reverse_translation": ["bordo", "consiglio", "pensione", "commissione", "tavola", "comitato"], "synset_id": [34656, 55801, 92], "score": 0.0188},
      {"word": "panel", "reverse_translation": ["pannello", "quadro", "gruppo", " riquadro", "tavola", "giuria"], "synset_id": [34656], "score": 0.0188},
      {"word": "plank", "reverse_translation": ["tavola", "asse", "tavolone", "pancone"], "synset_id": [92690], "score": 0.0040234118},
      {"word": "slab", "reverse_translation": ["lastra", "piastra", "pezzo", "tavola", "tavoletta", "tavolo anatomico"], "score": 0.00025720813},
      {"word": "diner", "reverse_translation": ["commensale", "trattoria", "tavola", "cliente"], "score": 8.7510656e-05}
    ]
  },
  "base_form": "tavola", "pos_enum": 1
},
"src": "it", "server_time": 5}

```

Figure 2: Google Translate response for the Italian word “tavola” translated into English

of the machine translation quality is BLEU (Bilingual Evaluation Understudy) (Papineni, Roukos, Ward, & Zhu, 2002), which is based on the n-gram precision model. Thus, this measure does not fit the context of word-to-word translation, the case that we are considering. A comparison between different machine translation tools is out of scope of this study. For a rich and comprehensive comparison of different machine translation tools we refer to the work of Costa-jussá, Farrús, Marino, and Fonollosa (2012).

BabelNet is arguably the largest state-of-the-art multilingual knowledge resource. BabelNet has integrated several Web-based collaborative multilingual knowledge resources (see Section 3.3). In addition, it makes different translation strategies available, which we want to evaluate indirectly in our study: sense-tagged sentence translation, direct machine translation of monosemous words, and translations obtained from Wikipedia to Wordnet mappings.

We used Google Translate and BabelNet to construct bilingual dictionaries for every pairs of non-English and English languages considered in our study (see Section 6.2).

Google Translate service is accessible through an API that can be used to translate *sentence-to-sentence* and *word-to-word* for many pairs of languages. Figure 2 shows Google’s word-to-word translation response in JSON (2015) format for the Italian word “tavola” translated into English. Google returns the preferable (common) translation in the **trans** item. A list of possible translations is also given in the **dict** item, which is part-of-speech (*PoS*) tagged. Each translation word in the **dict** item has a **reverse translation** set and a **score**. The **reverse translation** is a set of potential synonym words for the input word. The **score** estimates the translation usage (e.g., common, uncommon, or rare translations).

The translation directions (e.g., It-to-En, and En-to-It) of machine translation tools are said to have different performance when applied in cross-lingual information retrieval tasks (McCarley, 1999). To ensure the largest possible coverage we compiled three bilingual dictionaries using Google Translate by taking into account the translation direction. We also collect translations provided in the **reverse translation** sets. To the best of our knowledge available matching systems consider translations returned only in the **trans** item. For each pair of non-English and English languages considered in our gold standard we build the following bilingual dictionaries: *MT_fromEn* uses the translations collected from English to non-English words; *MT_toEn* uses the translations collected from non-English to English words; *MT* merges translations collected for the other two dictionaries to ensure the largest possible coverage (with Google Translate). Observe that *MT_fromEn* and *MT_toEn* are subsets of *MT*.

Table 1: Translation settings

Bilingual Dictionary	Description
<i>MT_fromEn</i>	translations from English to non-English words using Google Translate
<i>MT_toEn</i>	translations from non-English to English words using Google Translate
<i>MT</i>	the union of <i>MT_fromEn</i> and <i>MT_toEn</i>
<i>BN</i>	all translations encoded in BabelNet except translation from Open Multilingual WordNet
<i>BN_core</i>	BabelNet core synsets translations
<i>MT&BN_core</i>	the union of <i>MT</i> and <i>BN_core</i>
<i>MT&BN</i>	the union of <i>MT</i> and <i>BN</i>

BabelNet is structured as a graph of nodes. Nodes, called *BabelNet synsets*, represent concepts or named entities, which are lexicalized in several languages. For instance, the Italian lexicalizations in a node represent an Italian synset, which represents an equivalent synset to its corresponding English lexicalization, which is synset in the English WordNet. The translation of a given word in a source language (e.g., It) into a target language (e.g., En) with BabelNet is given by every word in the target language, which localizes the same nodes that are lexicalized with the input word. For example, the Italian word “tavola” is lexicalization of 15 nodes¹⁰ (14 concepts, and 1 named entity). These nodes provide 25 possible translations (lexicalization) in English: {board, correlation table, place setting, plank, setting, table, tablet, gang plank, wood plank, plate, table setting, stretcher bar, Panel, Panel cartoon, Oil on panel, ..., etc}. Each word in this lexicalization may derive from one or many of the different lexical resources integrated in BabelNet.

To analyze the impact of the different lexical resources integrated in BabelNet, we extracted, for every pair of non-English and English languages used in our study (see Section 6.2), two bilingual dictionaries from the BabelNet synsets. A first dictionary is extracted from BabelNet *core* synsets (called *BN_core*), which contain multilingual lexicalizations built from: sense-tagged sentences, monosemous word translation using Google Translate (monosemous words heuristic), and Wikipedia inter-lingual links. A second dictionary is extracted from BabelNet synsets (called *BN*), all synsets in BabelNet, which contain multilingual lexicalizations built from: *BN_core*, and lexicalization obtained from WikiData, Wiktionary, OmegaWiki, and Wikipedia redirection links. Observe that *BN_core* is a subset of *BN*. We only excluded translations obtained from the Open Multilingual WordNet (Bond & Foster, 2013), which we adopt as gold standards in our study.

We also merged translations from *BN_core* and *MT* dictionaries (called *MT&BN_core*), and translations from *BN* and *MT* dictionaries (called *MT&BN*). In this way we can compare and evaluate the impact of different Web-based multilingual resources, BabelNet core synsets, and the machine translation tools on the cross-lingual mapping tasks. The bilingual dictionaries we use in this study are summarized in Table 1.

6. Experiments

Three experiments are conducted to study the coverage, correctness, and impact of two multilingual lexical resources that we used as sources of translation on mapping concepts lexicalized in different languages. Four non-English wordnets, which are mapped to the English WordNet, are used as our gold standards.

10. <http://babelnet.org/search?word=tavola&lang=IT>

Table 2: Size of the wordnets (gold standards) used in the experiments

	English	Arabic	Italian	Slovene	Spanish
Words	147306	13866	40178	39985	36880
Word senses	206941	23481	61588	70947	57989
Synsets	117659	10349	33731	42583	38702

First, in Section 6.2 we describe in details the wordnets and we profile their concepts based on their lexicalizations. Then, in Section 6.2, we move to perform our experiments. We organize the discussion about our experiments as follows. In Section 6.3.1 we evaluate the coverage and correctness of translations obtained with different lexical resources to discuss their impact on retrieving candidate matches in concept mapping tasks. In Section 6.3.2, evidence collected from translations is used in a baseline mapping selection approach, i.e., majority voting, to evaluate the difficulty of the mapping selection task. In Section 6.3.3, we analyze the coverage of translations in relation to the position of the concepts in the semantic hierarchies.

Finally, in Section 6.4 we summarize our observations and draw some potential future directions.

6.1 Experimental Setup

The English, Arabic, Italian, Slovene, and Spanish wordnets are imported into a database. The wordnets database includes the words, synsets, semantic relations, and the mappings between each non-English wordnet and the English Wordnet. We compiled different bilingual dictionaries (Table 1) with Google Translate API and BabelNet as described in Section 5. We stored the dictionaries in a database so as to efficiently execute our experiments.

6.2 Mapped Wordnets Used as Gold Standards

In our study we use wordnets for English (Miller, 1995; Fellbaum, 1998), Arabic (Rodríguez et al., 2008), Italian (Pianta et al., 2002), Slovene (Fišer, 2007) and Spanish (Gonzalez-Agirre et al., 2012). These wordnets provide high quality cross-lingual mappings and contain very large inventories of concepts. Their size in terms of words, word senses and synsets is reported in Table 2¹¹. These wordnets have been built using different approaches and cover different families of languages: the *Germanic* languages (e.g., English), the *Romance* languages (e.g., Italian and Spanish), the *Slavic* languages (e.g., Slovene), and the *Semitic* languages (e.g., Arabic). Spanish, English, and Arabic are also among the top five spoken languages in the world (Wikipedia, 2015a), and their processing has gathered significant interest from the research community. Italian and Slovene represent two minority languages.

In Table 3 we show the distribution of words in each wordnet disaggregated by several categories: By considering *word ambiguity*, we distinguish between *Monosemous* words (M), words that have only one sense (meaning), and *Polysemous* words (P), words that have two or more senses. By considering *word complexity*, we distinguish between *Single* words (S), strings (lexemes) that have no spaces or hyphens, or *Collection* words (C), strings that consist of two or more simple words, which are connected by spaces or hyphens. We also

11. The Arabic, Italian, and Slovene wordnets are obtained from OMWN (2015), and the Spanish wordnet is obtained from MCR (2012). All *lexical gaps* (synsets with no lexicalization) (Vossen, 2004) are excluded.

Table 3: Word distribution in the gold standards by category: quantity (percentage)

Words	English	Arabic	Italian	Slovene	Spanish
<i>Monosemous(M)</i>	120433 (81.8)	10025 (72.3)	29816 (74.2)	28635 (71.6)	30106 (81.6)
<i>Polysemous(P)</i>	26873 (18.2)	3841 (27.7)	10362 (25.8)	11350 (28.4)	6774 (18.4)
<i>Simple(S)</i>	83118 (56.4)	8953 (64.6)	33133 (82.5)	29943 (74.9)	22630 (61.4)
<i>Collection(C)</i>	64188 (43.6)	4913 (35.4)	7045 (17.5)	10042 (25.1)	14250 (38.6)
<i>M&S</i>	59021 (40.1)	5361 (38.5)	22987 (57.2)	19223 (48.1)	16212 (44.0)
<i>M&C</i>	61412 (41.6)	4664 (33.6)	6827 (17.0)	9412 (23.5)	13894 (37.7)
<i>P&S</i>	24097 (16.4)	3592 (26.0)	10146 (25.3)	10720 (26.8)	6418 (17.4)
<i>P&C</i>	2776 (01.9)	249 (01.8)	218 (00.5)	630 (01.6)	356 (00.9)

Table 4: Synsets categories

Category	Synset name	Definition “synsets that have...”
<i>all_M</i>	<i>all words Monosemous</i>	only monosemous words
<i>all_P</i>	<i>all words Polysemous</i>	only polysemous words
<i>OWS</i>	<i>One-Word</i>	only one word (<i>synonymless synset</i>)
<i>MWS</i>	<i>Many-Words</i>	two or more synonym words (<i>synonymful synset</i>)
<i>M&OWS</i>	<i>Monosemous and OWS</i>	only one word, which is also a monosemous word
<i>M&MWS</i>	<i>Monosemous and MWS</i>	two or more synonym words, which are all monosemous words
<i>MIX</i>	<i>MIXed</i>	monosemous and polysemous synonym words
<i>P&OWS</i>	<i>Polysemous and OWS</i>	only one word, which is also a polysemous word
<i>P&MWS</i>	<i>Polysemous and MWS</i>	two or more synonym words, which are polysemous words

consider the four categories that are derived by combining word ambiguity and complexity categories. For example, “*tourism*” is a monosemous and simple word (*M&S*), “*tabular array*” is a monosemous and collection word (*M&C*), “*table⁺*” is a polysmouse and simple word (*P&S*), and “*break up⁺*” is a polysemous and collection word (*P&C*).

Observation 1. A vast majority of collection words are monosemous words: only an average of 1.3% words are polysemous collection words across all wordnets. This means that a word used as concept label is less likely to be ambiguous if it is a composite word and more likely to be ambiguous if it is a simple word.

We can classify the synsets based on the ambiguity and number of their words (respectively first and second, and third and fourth categories of synsets described in the upper part of Table 4). By combining these orthogonal classifications, we can consider five categories of synsets as described in the lower part of Table 4. One can observe that the *M&OWS* and the *M&MWS* are subsets of the *all_M*. The *P&OWS* and the *P&MWS* are subsets of the *all_P*, and the *MIX* are subsets of the *MWS*. Examples of synsets in the English WordNet for each category are shown in Table 5. Table 6 describes, for every wordnet, the total number and percentage of synsets grouped by category.

Table 5: Synset examples for all categories in English

Category	Example	Definition
<i>M&OWS</i>	{desk}	a piece of furniture with a writing surface and usually drawers or other compartments
<i>M&MWS</i>	{tourism, touristry}	the business of providing services to tourists
<i>MIX</i>	{table ⁺ , tabular array}	a set of data arranged in rows and columns
<i>P&OWS</i>	{cocktail ⁺ }	a short mixed drink
<i>P&OWS</i>	{cocktail ⁺ }	an appetizer served as a first course at a meal
<i>P&OWS</i>	{table ⁺ }	a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs
<i>P&MWS</i>	{board ⁺ , table ⁺ }	food or meals in general

Table 6: Synset category-wise distribution in gold standards: quantity (percentage)

Synsets	English	Arabic	Italian	Slovene	Spanish
all_M	57415 (48.8)	3381 (32.7)	14393 (42.7)	17615 (41.4)	19020 (49.1)
all_P	41568 (35.3)	4409 (42.6)	14641 (43.4)	19609 (46.0)	16269 (42.1)
MWS	53784 (45.7)	6162 (59.5)	13644 (40.4)	14994 (35.2)	14994 (38.7)
OVS	63875 (54.3)	4197 (40.5)	21084 (59.6)	27589 (64.8)	27589 (71.3)
M&OVS	33596 (28.6)	1995 (19.3)	10492 (31.1)	14848 (34.9)	14120 (36.5)
M&MWS	23819 (20.2)	1386 (13.4)	3901 (11.6)	2767 (06.5)	4900 (12.7)
MIX	18676 (15.9)	2559 (24.7)	5691 (16.9)	5359 (12.6)	3413 (08.8)
P&OVS	30279 (25.7)	2194 (21.2)	9609 (28.5)	12741 (29.9)	12005 (31.0)
P&MWS	11289 (9.60)	2215 (21.4)	4046 (12.0)	6868 (16.1)	4264 (11.0)

Table 7: Examples of mappings between Italian and English synsets by category

Synsets	M&OVS	M&MWS	MIX	P&OVS	P&MWS
M&OVS	{scuola d'arte, {art school}}	{radiostazione, {radio station}}	{tavolino, {desk}, banco ⁺ , scrivania}	{ordinario ⁺ , {full professor}}	{entita' ⁺ , {entity}, cosa ⁺ }
M&MWS	{turismo, {tourism, touristry}}	{accoppiata, {exact, abbinata}, {perfecta}}	{docente ⁺ , {prof, cattedratico, professore}}	{viaggiatore ⁺ , {traveler, traveller}}	{classe ⁺ , {classroom, schoolroom}}
MIX	{minorit ⁺ , {minority, nonage}}	{biforcarsi, {branch, ramificarsi, furcate, diramarsi}, {separate}}	{tavola ⁺ , {table, tabular array}}	{contribuire ⁺ , {conduce, contribute, lead ⁺ }}	{cibo ⁺ , {repast, pasto ⁺ , meal ⁺ , mangiare ⁺ }}
P&OVS	{forchetta, {fork ⁺ }}	{stretto, {close ⁺ , vicino}}	{poltrona ⁺ , {chair ⁺ , seggiola, sedia}}	{cosa ⁺ , {thing ⁺ }}	{tavola ⁺ , {table ⁺ , tavolo ⁺ }}
P&MWS	{chiudersi, {close ⁺ , shut ⁺ }}	{inquietarsi, {care ⁺ , allarmarsi}, {worry ⁺ }}	{segnare ⁺ , {score ⁺ , scalfire, mark ⁺ , nock ⁺ }}	{moderare ⁺ , {chair ⁺ , moderate ⁺ , lead ⁺ }}	{cibo ⁺ , {board ⁺ , vitto ⁺ , table ⁺ }}

Observation 2. Wordnets have more synonymless synsets (*OVS*) than synonymful synsets (*MWS*), with 58.1% of synsets being, on average, synonymless. Arabic, which has less *OVS* than *MWS*, represents an exception among the considered wordnets. In particular, the Arabic polysemous synsets (*all_P*) are equally distributed between *OVS* and *MWS*.

In our gold standards there exist mappings between synsets of every category. Examples of mappings for each couple of categories of synsets from Italian to English are shown in Table 7. The percentage of the mapped synsets between the non-English wordnets and the English WordNet, grouped by category, is reported in Table 8.

The results confirm that languages do not cover the same number of words as noticed by Hirst (2004), and, hence, concepts shared in different languages have different ways to express their meanings (i.e., they belong to different lexical categories). For instance, 57% of the Italian *M&OVS* synsets are mapped to *monosemous* synsets in English (*M&OVS* and *M&MWS*). On the other hand, 25% of the Italian *M&OVS* are mapped to *polysemous* synsets in English (*P&OVS* and *P&MWS*). The percentage of monosemous non-English synsets that are mapped to the polysemous English synsets ranges from 10% (Slovene) to 30% (Arabic). The percentage of the monosemous English synsets that are mapped to the polysemous non-English synsets ranges from 6% (Arabic) to 14% (Italian). For instance, the *M&OVS* Italian synsets {fotografare} and {azioni ordinarie} are mapped to {shoot⁺, snap⁺, photograph⁺} and {common shares, common stock, ordinary shares}, respectively a *P&MWS* and an *M&MWS* English synset.

Table 8: Distribution of mapping by category: percentage

English	M&OWS	M&MWS	MIX	P&OWS	P&MWS	M&OWS	M&MWS	MIX	P&OWS	P&MWS
	Arabic					Italian				
M&OWS	32.9	19.2	5.1	5.4	2.3	36.2	20.9	10.6	9.4	4.1
M&MWS	15.1	28.6	5.1	2.5	1.5	21.2	34.9	10.3	4.6	2.8
MIX	17.2	28.7	37.7	15.5	22.6	17.8	27.2	38.7	22.5	26.8
P&OWS	27.4	14.8	21.7	57.3	29.5	17.9	10.7	18.4	43.0	29.0
P&MWS	7.3	8.7	30.4	19.4	44.2	6.9	6.3	22.0	20.5	37.4
English	Slovene					Spanish				
M&OWS	23.4	25.2	14.2	9.0	6.8	42.6	10.7	7.8	8.4	3.1
M&MWS	47.8	39.7	13.0	4.4	4.3	22.2	63.1	7.7	3.3	1.9
MIX	18.1	27.5	48.7	20.2	27.1	14.5	17.1	44.1	19.4	24.2
P&OWS	7.1	4.0	8.4	45.3	25.7	17.8	5.4	15.1	48.5	25.9
P&MWS	3.5	3.7	15.7	21.1	36.1	2.9	3.8	25.3	20.4	44.9

Observation 3. Synsets in different languages, which have an equivalent meaning, can fall in different synset categories. For example, the Italian monosemous synonymless synset {forchetta} is mapped to the polysemous synonymless synset {fork⁺} in English. This indicates that the monosemous word heuristic, which is adopted by some approaches to concept mapping and multilingual knowledge construction, e.g., work presented by Navigli and Ponzetto (2012), is successful for a large number of concepts but fails for still a relevant number of concepts. An average of 19.3% non-English monosemous synsets are mapped to English polysemous synsets in the gold standards, and an average of 8.9% English monosemous synsets are mapped to non-English polysemous synsets in the gold standards. More details on the impact of the monosemous word heuristics are provided in Section 6.3.1, where translation correctness is analyzed.

6.3 Results and Discussion

In this section we describe in details the three experiments presented in our study.

6.3.1 EXPERIMENT 1: COVERAGE AND CORRECTNESS OF TRANSLATIONS FOR CANDIDATE MATCH RETRIEVAL

In order to evaluate the **coverage** of the translations obtained with different lexical resources, we use two measures. We compute the *average word sense coverage* across all words of a wordnet, where word sense coverage is defined for an individual word as in Eq.6. We compute the *average synset coverage* across all synsets of a wordnet, where synset coverage is (a boolean value) defined for an individual word as in Eq. 7. All values are normalized in the range [0..100]. For sake of clarity we will simply refer to these measures as word sense and synset coverage (at the wordnet level).

Table 9 reports, for each wordnet, word sense and synset coverage with different translation settings. Synsets have higher coverage than word senses in all the translation settings. This can be explained with the observation that a synset is covered if its translation returns at least one word of the lexicalization of its equivalent synset in the target language (see Eq.7).

We observe that machine translations from non-English *to English* (*MT_toEn*) achieve higher word sense and synset coverage than machine translation *from English* to non-English (*MT_fromEn*). For instance, word sense coverage of *MT_toEn* is from 5.2 (Italian) to 10.9

Table 9: Word sense and synset coverages with different translation settings

Translation	Arabic		Italian		Slovene		Spanish	
	Senses	Synsets	Senses	Synsets	Senses	Synsets	Senses	Synsets
<i>BN_{core}</i>	19.9	37.4	40.0	62.5	28.8	44.2	33.9	44.7
<i>BN</i>	30.8	51.3	51.7	72.8	35.9	52.0	39.8	49.0
<i>MT_{fromEn}</i>	51.3	69.9	60.2	81.9	40.2	60.0	56.1	67.8
<i>MT_{toEn}</i>	57.9	76.1	65.4	83.9	49.6	67.2	67.0	77.0
<i>MT</i>	59.2	77.7	68.1	87.6	53.8	72.4	69.4	79.7
<i>MT&BN_{core}</i>	60.8	79.2	69.8	89.0	55.8	74.2	71.5	81.3
<i>MT&BN</i>	62.5	80.2	72.2	89.9	57.5	75.2	72.3	81.7

(Spanish) percentage points higher than *MT_{fromEn}*, and synset coverage of *MT_{toEn}* is from 2.0 (Italian) to 9.2 (Spanish) percentage points higher than *MT_{fromEn}*.

Observation 4. Machine translation tools perform asymmetrically: *MT_{toEn}* achieves higher word sense and synset coverage than *MT_{fromEn}*.

The machine translation bilingual dictionary (*MT*), which we built from the union of both machine translation directions (see Section 5), performs better than the dictionaries that we built considering each direction alone (i.e., *MT_{fromEn}* or *MT_{toEn}*). Word sense coverage of *MT* is on average 2.7 and 8.2 percentage points higher than *MT_{toEn}* and *MT_{fromEn}*, respectively. Synset coverage of *MT* is on average 3.5 and 7.0 percentage points higher than *MT_{toEn}* and *MT_{fromEn}*, respectively.

BN_{core} and *BN* translation settings, which are based on BabelNet, obtain lower coverage than every machine translation setting for all wordnets. This can be explained by limited coverage of the words that occur in non-English wordnets by Wikipedia concepts (which mostly cover named entities), and by incompleteness of the mappings used to construct BabelNet (Navigli & Ponzetto, 2012). However, it should be remarked that, for several languages, BabelNet also includes the lexicalizations from the Open Multilingual WordNet that have been excluded in our study because it is part of the gold standard (see Section 6.2). This means that for several well-known languages such as French, Germany, Spanish, and Italian¹² we can expect much higher translation coverage from BabelNet. Still, best results are obtained when combining all available translations, i.e., both from the machine translation tool and BabelNet, *MT&BN*. For instance, *MT&BN* word sense coverage is on average 3.5 percentage points higher than *MT*. *MT&BN* synset coverage is on average 2.4 percentage points higher than *MT*.

We also observe that *BN* achieves considerably higher coverage than *BN_{core}*, with an average difference in word sense and synset coverage of 10.4 and 10.1 percentage points respectively (*BN_{core}* is a subset of *BN* - see Section 5). However, most of this additional coverage is lost when combining *BN_{core}* and *BN* with *MT* translations: *MT&BN* word sense coverage is on average only 1.7 percentage points higher than *MT&BN_{core}*, and *MT&BN* synset coverage is on average 0.8 percentage points higher than *MT&BN_{core}*.

Observation 5. The results highlight that machine translation tools achieve higher coverage than BabelNet, which integrates several Web-based multilingual resources (i.e., Wikitionary, OmegaWiki, WikiData, and Wikipeida redirection links). However, integrating BabelNet with machine translation tools still yields a significant gain in coverage, mostly

12. See the Languages and Coverage Statistics at BabelNet (2012).

Table 10: Average synset coverage by category

Synsets	Arabic			Italian			Slovene			Spanish		
	BN	MT	MT&BN	BN	MT	MT&BN	BN	MT	MT&BN	BN	MT	MT&BN
<i>all_M</i>	35.4	64.2	67.3	68.6	86.0	88.4	59.8	78.8	80.8	34.4	78.1	79.5
<i>all_P</i>	58.8	82.9	85.4	69.0	80.5	83.0	44.8	65.5	69.1	62.5	80.2	83.0
<i>OWS</i>	44.5	67.0	70.5	64.1	79.5	82.2	52.8	70.1	73.0	44.1	75.8	78.0
<i>MWS</i>	56.1	85.0	86.8	81.0	93.6	95.2	50.7	76.6	79.1	59.0	87.7	89.3
<i>M&OWS</i>	32.2	58.0	61.5	63.8	83.5	86.1	60.8	78.3	80.3	32.4	74.7	76.2
<i>M&MWS</i>	40.3	73.4	76.0	81.6	92.6	94.5	54.3	81.4	83.2	40.1	88.0	89.0
<i>MIX</i>	59.8	86.9	88.6	80.7	94.3	95.8	53.1	76.5	79.0	65.5	85.8	87.6
<i>P&OWS</i>	55.8	75.4	79.0	71.0	83.3	86.4	43.4	60.5	64.5	57.9	77.1	80.2
<i>P&MWS</i>	61.9	90.5	92.0	80.8	93.7	95.1	47.3	74.7	77.6	75.5	88.8	90.9

Table 11: Average number of candidate matches

Synsets	Arabic	Italian	Slovene	Spanish
synonymless (<i>OWS</i>)	48	17	11	27
synonymful (<i>MWS</i>)	124	49	21	75

because of BN_{core} (Wikipedia inter-lingual links, and the context based translations).

Table 10 reports the average coverage for each synset category by using *BN*, *MT* and *MT&BN* translation settings (the settings achieving highest coverage). The results show that the *synonymful* synsets (*MWS*) are covered more than *synonymless* synsets (*OWS*) for every wordnet and almost every translation setting. This confirms the intuition that richer concept lexicalizations help to find at least one correct translation using machine translation tools. *Polysemous* synsets (*all_P*) are covered more than the *monosemous* synsets (*all_M*) for Arabic and Spanish, but less than *monosemous synsets* (*all_M*) for Italian and Slovene. This can be explained by the distribution of polysemous and monosemous synsets between synonymless and synonymful synsets: most of the monosemous synsets (*all_M*) are synonymless synsets, and most of the polysemous synsets (*all_P*) are synonymful synsets. *MIXed* synsets are the most covered synsets, since they are synonymful synsets, which combine monosemous and polysemous words.

Observation 6. Synonymful synsets (*MWS*) are covered more than synonymless synsets (*OWS*) (see Table 10). However, a higher coverage comes at the price of a larger number of candidate matches, thus making the mapping selection task more challenging (see Table 11).

Observation 6 can be supported by figures shown in Table 11, which reports the average number of candidate matches for synonymless vs. synonymful synsets. In addition, most of synonymful synsets contain at least one polysemous word (see Table 6). Thus, one can expect that the sets of candidate matches returned by translations of synonymful synsets are not only larger in size, but also noisier, because of the translation of polysemous words. A more in-depth analysis on the difficulty of the mapping selection task for the different synset categories is provided in Section 6.3.2. Such analysis will confirm that the mapping selection problem is more difficult for synsets that contain polysemous words, which represent the majority of synonymful synsets. At the same time, the joint translation of synonym words can support mapping selection for many synsets (e.g., for synsets that do not contain only

Table 12: Average recall and word-translation correctness by category

Words	Arabic						Italian					
	BN		MT		MT&BN		BN		MT		MT&BN	
M	20.2	(63.6)	45.1	(36.9)	48.0	(56.4)	49.0	(65.6)	65.8	(47.0)	69.8	(62.1)
P	53.1	(38.0)	83.3	(22.8)	85.2	(40.9)	71.5	(44.9)	89.8	(31.9)	91.3	(45.1)
S	38.1	(48.3)	67.0	(27.3)	70.0	(49.1)	54.4	(57.0)	73.0	(41.0)	75.9	(55.5)
C	13.3	(63.4)	35.1	(43.9)	37.0	(53.8)	56.9	(65.8)	67.3	(47.5)	72.6	(62.8)
M&S	26.7	(63.0)	54.8	(32.6)	58.6	(57.2)	46.7	(65.3)	65.4	(46.7)	69.1	(61.6)
M&C	12.7	(65.2)	34.0	(44.9)	35.8	(55.0)	56.9	(66.3)	66.9	(47.9)	72.2	(63.5)
P&S	55.0	(37.7)	85.2	(22.3)	87.0	(40.9)	71.8	(44.8)	90.0	(31.7)	91.4	(45.1)
P&C	25.3	(46.6)	55.4	(32.0)	58.2	(40.3)	56.0	(48.6)	80.3	(39.0)	84.9	(44.1)
ALL	29.3	(50.8)	55.7	(31.0)	58.3	(50.2)	54.8	(58.6)	72.0	(42.1)	75.3	(56.8)

Words	Slovene						Spanish					
	BN		MT		MT&BN		BN		MT		MT&BN	
M	45.2	(66.1)	63.6	(47.8)	66.6	(60.8)	28.1	(61.6)	68.4	(48.1)	70.8	(56.9)
P	42.6	(39.6)	73.0	(30.1)	75.4	(33.7)	74.8	(38.9)	92.1	(28.4)	93.7	(41.4)
S	43.7	(56.3)	67.0	(41.4)	69.7	(49.9)	48.9	(51.0)	78.8	(41.0)	81.0	(53.3)
C	46.7	(66.0)	64.3	(45.0)	67.3	(60.3)	17.4	(62.3)	63.1	(48.4)	65.7	(53.5)
M&S	43.7	(65.8)	62.5	(49.0)	65.4	(60.7)	37.6	(61.1)	73.2	(47.6)	75.5	(59.1)
M&C	48.3	(66.5)	66.0	(45.5)	69.1	(61.1)	17.0	(63.0)	62.8	(48.6)	65.4	(53.9)
P&S	43.7	(39.2)	75.0	(30.0)	77.4	(33.5)	77.3	(38.7)	93.0	(27.9)	94.7	(41.5)
P&C	22.9	(50.3)	38.9	(31.3)	41.1	(39.9)	30.6	(47.5)	75.6	(38.8)	76.7	(40.3)
ALL	44.5	(58.9)	66.3	(42.3)	69.1	(52.4)	36.7	(53.1)	72.8	(43.5)	75.0	(53.3)

polysemous words, e.g., *MIXed* synsets), as a means to collect evidence for deciding about a mapping.

In order to evaluate the **correctness** of the translations obtained with different resources, we use two measures. We compute the *average word-translation correctness* across all words of a wordnet; word-translation correctness is defined for an individual word as in Eq.5. In addition, we report *average word-translation recall* (*recall*, for short), using the subformula in Eq.5¹³.

Average recall and word-translation correctness for the *BN*, *MT* and *MT&BN* dictionaries, disaggregated by word category, are reported in Table 12.

The results show that word-translation is more correct for monosemous and collection words than for polysemous and simple words. In contrast, recall of word-translation is higher for *polysemous* words (*P*) than for *monosemous* words (*M*) with every source of translation and for every wordnets, with the exception of the *BN* dictionary for the Slovene wordnet. Recall of word-translation is also higher for *simple* words (*S*) than for *collection* words (*C*) in every setting. These observations can be explained by monosemous words being usually less frequent and more domain-specific than polysemous words. In addition, most of collection words are also monosemous words - as remarked in *Observation 1* -, while most of polysemous words are simple words: recall and correctness of translations for simple words is affected by the translation of simple polysemous words. Translations of polysemous and simple words return on average larger word sets. These sets are more likely to contain richer lexicalizations in the target language, but also to contain words that do not belong to any sense of the input words in the target language.

Observation 7. The translation of monosemous and collection words is on average more correct than the translation of polysemous and simple words, but achieves lower recall.

13. Word-translation correctness is defined using a formula based on F_1 -measure.

Focusing on the performance of different sources of translation, we can notice that recall for *MT* is higher than for *BN*, while correctness for *BN* is higher than for *MT*. *MT&BN* combines the strengths of both dictionaries, i.e., higher recall of word-translation because of *MT*, and higher correctness because of *BN*. For instance, in Table 12 we can notice that the correctness of word-translation is improved by 9.8 and 19.2 percentage points for Spanish and Arabic respectively, if we add to *MT* translations derived from *BN*. The recall of word-translation is improved by as much as 20.4 and 38.3 percentage points for Italian and Spanish respectively, if we add to *BN* translations derived from *MT*. The best results are thus obtained for *MT&BN*, for which we obtain recall (correctness) scores that range from 58.3%(50.2%) (Arabic) to 75.3% (56.8%)(Italian). The low recall for Arabic can be explained by a low recall for translations of monosemous collection words.

Observation 8. The combination of machine translation tools with Web-based multilingual resources and context-based sentence translations, like the ones incorporated in BabelNet, improves not only the recall, but also the correctness of word-translations.

6.3.2 EXPERIMENT 2: MAPPING SELECTION DIFFICULTY

On one hand, the translations returned for a given synset can be used as evidence to select a mapping to a synset in a target language. On the other hand, translations of many, polysemous words in a synset can return several candidate matches, most of which are incorrect, thus making the mapping selection task difficult to solve. This experiment analyzes the difficulty of the mapping selection task when performed over candidate matches retrieved with translations obtained from different lexical resources.

In this experiment we use translations returned by the *MT* machine translation tool, for the sake of simplicity (with the exception of the analysis of the synonym word coverage, where we also include *BN*). We focus on *MT* because, as shown in the previous section, it has higher coverage than *BN*, and it has been widely used in previous work on ontology matching. In addition, the slight increase in coverage obtained with *MT&BN*, when compared to *MT*, can be ignored for this particular experiment.

To perform our analysis we use a greedy baseline method for candidate mapping selection and we compare the quality of the alignment computed with this method to the gold standard alignment. As a baseline mapping selection method we use *majority voting* on top of the evidence collected from synset translations.

Mapping Selection with Majority Voting. Every source synset is translated using the synset translation function defined in Eq.2. The output is represented as the multi set union of the returned translations. Each word $w^{(i)}$ in the multi set, with (i) being the word frequency count, represents i votes for all the candidate matches that contain w . Therefore, a candidate match t for a source synset s , such that t contains many words returned by the translation of s , will receive more votes and will be more likely to be the target of the selected mapping. Candidate matches are ranked by votes and the mapping containing the top-voted match is selected.

It can happen that several candidate matches receive an equal number of votes, which results in a *tie*. In this case, for a source synset the mapping selection task is *undecidable*; in contrast we will say that a mapping is *decidable* when a unique candidate match receives the highest number of votes. However, when a tie occurs among a set of top-voted candidate

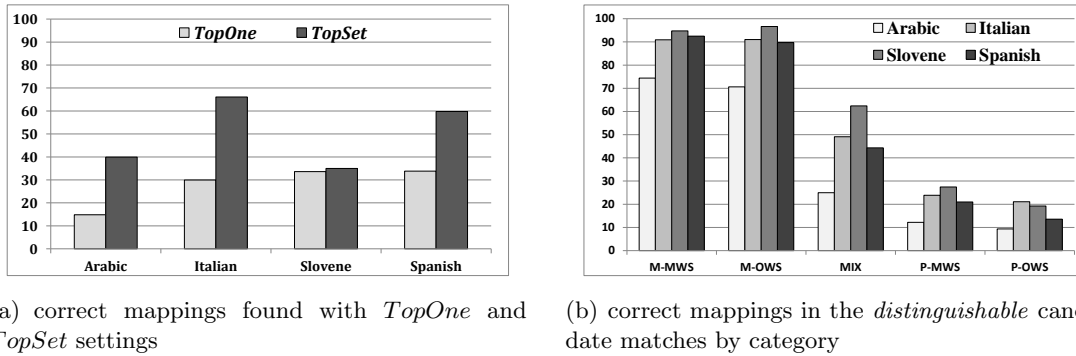


Figure 3: Correct mappings found with baseline selection strategy

matches, it is valuable to know if this set contains also the correct mapping (w.r.t to the gold standard) and the number of candidate matches in the tie. In fact, if the set of top-voted candidate matches also contains the correct match, the correct mapping could be found via user interaction with relatively low effort. For these reasons we use two settings in the experiments with a majority voting candidate selection approach:

- **TopOne:** if there exists a unique top-voted candidate match for a source synset, the mapping containing this match is selected and included in the alignment. If a tie occurs, no mapping is selected.
- **TopSet:** the correct mapping is selected by an oracle from the set of top-voted matches (no matter of its cardinality) and included in the alignment.

To quantify the quality of the alignment we compute (*selection*) *correctness* as the percentage of the correct mappings returned by each selection setting *over the set of covered mappings*, i.e., the mappings for which the set of candidate matches contains the correct mapping¹⁴. In other words, in *TopOne* setting, a mapping is considered correct for a source synset, only when the correct match for the synset (according to the gold standard) is its unique top-voted candidate match; in *TopSet* setting, a mapping is considered correct for a source synset, whenever the correct match for the synset is included in the set of its top-voted candidate matches. Observe that every mapping that is counted as correct in *TopOne* setting, is also counted as correct in *TopSet* setting.

A comparison between the performance in terms of correct mappings returned in the *TopOne* and *TopSet* selection settings for each wordnet is shown in Figure 3(a). The average of correct mappings obtained in *TopOne* and *TopSet* settings is 28% and 50% respectively. Based on the performance of such simple baseline methods, we suggest that translations can be helpful for mapping selection, although more sophisticated methods to make use of their evidence have to be devised. In addition, number of correct mappings can be increased up to an average of 30 points in the case where we assume that a user can select the correct mapping among the set of top-voted matches returned by a mapping selection method,

14. This is equivalent to compute a *relative precision* measure: Precision is interpreted as usual in ontology matching (Shvaiko & Euzenat, 2013) but normalized in the range [0..100], and evaluated only over a restricted subset of the gold standard. Such a restricted subset consists of all the mappings containing source concepts that are covered by translations

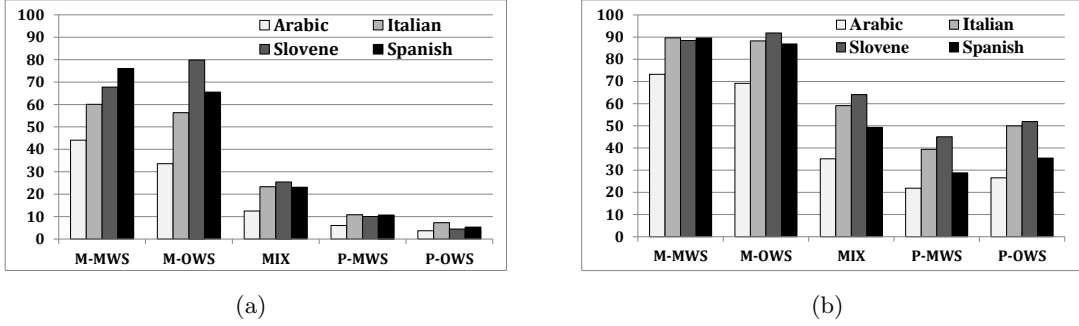


Figure 4: Percentage of the correct mappings by synset category with (a) *TopOne* selection and (b) *TopSet* selection

e.g., with an interactive ontology matching approach (Cruz, Loprete, Palmonari, Stroe, & Taheri, 2014; Sarasua, Simperl, & Noy, 2012). However, the average cardinality of the sets of top-voted matches (*TopSet*) is as high as 49 synsets, which makes it difficult for users to make a decision.

Figure 3(b) shows, for every wordnet and every category of source synset, the percentage of correct mappings found using *TopOne* selection over the total of synset with decidable mappings. The baseline *TopOne* mapping selection strategy achieves a remarkable performance for monosemous synsets (i.e., *M&OWS* and *M&MWS*) and poor performance for polysemous synsets. On average, *TopOne* selection is capable to select correct matches for as much as 88.2% of the monosemous synsets.

Figure 4(a) and 4(b) show, for every wordnet and every category of target synset, the percentage of correct mappings found respectively with *TopOne* and *TopSet* selection settings. We figured out that mappings to synsets with polysemous words, in particular to *polysemous synonymless synsets* (*P&OWS*), are much more likely to be undecidable, i.e., a set of many top-voted candidate matches is found. In fact, when the target synsets are *P&OWS*, the mapping is almost always undecidable with the *TopOne* selection.

Observation 8. Evidence provided by machine translation tools is valuable to successfully decide upon correct mappings for monosemous synsets, while it fails to support such a decision for most of the polysemous synsets.

Observation 9. Mappings with polysemous and synonymless target synsets (*P&OWS*) cannot be successfully selected by leveraging only the evidence from translations and a simple selection strategy like majority voting because translations assign an equal number of votes to several candidate matches.

Observation 10. If the set of top-voted candidate matches can be validated, e.g., as in the *TopSet* selection settings, it is possible to find a correct mapping for a vast majority of monosemous synsets (on average, 85%).

We want to investigate if correct mappings are more likely to be found when a larger or a small number of top-voted mappings is selected (with *TopSet* selection). To this end, we analyze the distribution of the correct mappings found with *TopSet* selection among top-voted candidate matches of different size for every wordnet. Correct mappings are found

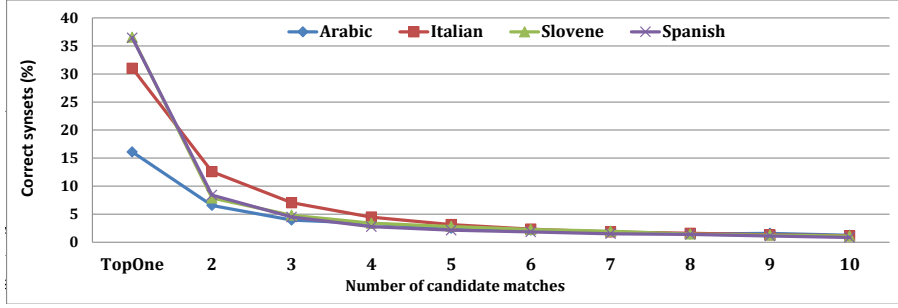


Figure 5: Percentage of correct mappings vs. size of top-voted candidate matches with *TopSet* selection

Table 13: Synonym words coverage (%) for synonymful synsets (MWS)

Translation	Arabic	Italian	Slovene	Spanish
BN	51.9	59.8	56.7	61.2
MT	68.9	68.5	61.5	74.4
MT&BN	71.3	72.6	65.4	77.5

in sets of top-voted candidate matches with a size that ranges from 1 to 238 candidates. The distribution is plotted in Figure 5: *x-axis* represents the number of selected top-voted candidate matches (up to size equal to ten), while the *y-axis* represents the percentage of found correct mappings. On average, 28% of correct mappings are found when a unique top-voted candidate match exists, i.e., like in *TopOne* selection settings (see Figure 3(a)). For instance, about 4% of the correct mappings are found in sets of top-voted mappings that contain four candidate matches, a percentage that represents an absolute number of 317, 1455, 991, and 1328 synsets for the Arabic, Italian, Slovene, and Spanish wordnets, respectively.

Observation 11. Synsets that occur as targets in mappings found with *TopOne* selection (decidable mappings) can be safely filtered out from candidate matches for other source synsets, with an error estimated to be as low as 0.2% of removing a correct match.

Finally, we analyze the impact of synonyms on the mapping selection task. Synonymful synsets (i.e., *M&MWS*, *MIX*, and *P&MWS*) are more likely to be correctly mapped with *TopOne* selection (Figure 4(a)) than synonymless synsets (i.e., *M&OWS* and *P&OWS*), even if the average number of candidate matches is greater for synonymful synsets than for synonymless synsets (see Table 11). These results confirm that synonyms are helpful not only for retrieving candidate matches - as previously observed in *Observation 6* - but also for selecting the correct mappings: the translation of different words that express a same concept provide evidence to decide the best mapping for this concept.

Table 13 reports, for every wordnet, the synonym words coverage for synonymful synsets (*MWS*) using the *BN*, *MT* and *MT&BN* dictionaries (synonym words coverage is defined by Eq.8). The best results are obtained with *MT&BN*, with synonym words coverage ranging from 65.4% (Slovene) to 77.5% (Spanish). Thus, on average, more than two synonyms are translated correctly in synonymful synsets.

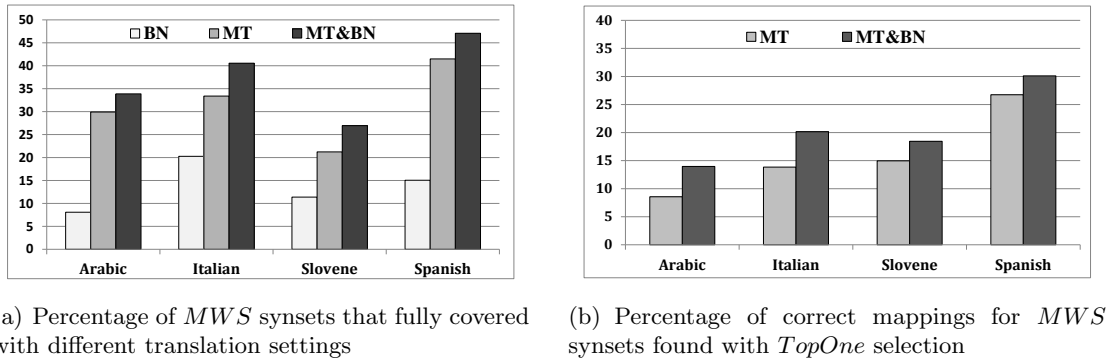

 Figure 6: Synonymful synsets (*MWS*) whose synonym words are fully covered

Figure 6(a) shows the percentage of synonymful synsets that are **fully covered**, i.e., synsets that contain only words that are correctly translated. On average, the *MT* dictionary fully covers a greater percentage of synonymful synsets than *BN*, with a gain of 18 points. The best results are obtained by *MT&BN* with an average gain of 6 points with respect to *MT*. Although the *BN* dictionary has limited impact on overall synsets coverage (with a gain of 2.4 points, as shown in *Experiment 2*), *BN* improves synonym words coverage by an average of 6 points, which can have a significant impact on mapping selection with majority voting. For instance, when compared to *MT*, the *MT&BN* dictionary improves the percentage of correct mappings in *TopOne* selection for the synonymful synsets that are fully covered by 4.6 points, as shown in Figure 6(b). Covering more synonym words belonging to a synonymful synset, not only improves synsets coverage, but also makes the mapping selection step easier. Thus, integrating more lexical resources for translation can be advantageous in the mapping selection tasks as well.

Observation 12. For synonymful synsets, the larger the number of synonym words covered by translations, the easier the mapping selection task is.

6.3.3 EXPERIMENT 3: COVERAGE AND CORRECTNESS OF TRANSLATIONS VS. CONCEPT SPECIALIZATION

We recall that a synset is not covered when none of the words of its equivalent synset in the target language is returned by its translation. In other words, when a synset is not covered, the correct match cannot be found among the set of candidate matches found by translation. This analysis further helps in the exploration of the problem of synset coverage by investigating 1) the impact of domain specificity on synset coverage, and 2) the possibility of improving the coverage by expanding the set of found candidate matches with synsets similar to the ones retrieved by translations.

To investigate if non covered synsets can be characterized to some extent based on their specificity, we use two different methods to characterize specificity: the domain labels associated with synsets in *WordNet Domains* (Bentivogli, Forner, Magnini, & Pianta, 2004), e.g., biology, animals, and so on; the position that synsets occupy in the semantic hierarchies, e.g., synsets that occur as leaf nodes in the hypernym hierarchies.

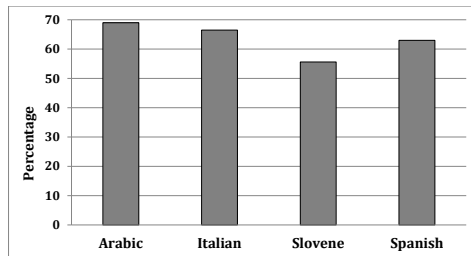


Figure 7: Percentage of *domain specific* synsets that are not covered by *MT*

We consider a synset which is associated with a domain label in Wordnet domains as domain-specific, i.e., every label different from “*Factoum*” (i.e., general, or non specified domain). For every wordnet, the percentage of domain specific synsets that are not covered by *MT* dictionary is shown in Figure 7. For example, we found that, on average, only 36% of the non covered synsets with the *MT* dictionary are labeled as “*Factoum*”. The rest of the non covered synsets (64%) are distributed over different domains (with biology, animals, person, plants, and geography as the most frequent ones). These findings consolidate the ones discussed in *Experiment 1*: monosemous words, which do often express specific concepts, were found to be less covered than polysemous words, which often express more general concepts.

Observation 13. Domain-specific concepts have less coverage, by machine translation tools, than general concepts.

With the same intent, we consider how synsets not covered by translations are distributed in the semantic hierarchy defined by the hypernym/hyponym relation. In this context, we consider leaf synsets (called *Lsynsets*) as the most specific synsets, while intermediate synsets (called *Isynsets*), i.e., synsets occurring in other positions in the hierarchy, are considered to be more generic. We consider only a subset of synsets, i.e., *nominal synsets*, whose hierarchical structure is well-established in the English wordnet. In particular, to determine the position of a source synset we consider the position of its equivalent synset in the English WordNet, by using the mappings existing between the wordnets.

Figure 8(a) reports the percentage of *Lsynsets* and *Isynsets* for every wordnet. We can notice that most of the wordnets have more leaf synsets than intermediate synsets, with the exception of the Arabic wordnet. This exception can be explained by the strategy used for the construction of this wordnet and by its relatively small size. The construction of the Arabic wordnet (Rodríguez et al., 2008), which is based on the expand model paradigm introduced in the EuroWordNet project (Vossen, 2004), was initiated by the translation of the *core concepts* of the English WordNet (Boyd-Graber, Osherson, & Schapire, 2006), and was, thereafter, extended to other concepts. The core concepts (over 5 thousands) are often assumed to be common across different cultures and languages, and are often intermediate synsets.

Figure 8(b) reports the percentage of *Lsynsets* and *Isynsets* that are not covered with *MT* dictionary for each wordnet. The average percentage of nominal *Lsynsets* and *Isynsets* not covered with the *MT* dictionary is 21.1% and 16.6%, respectively. Table 14 reports, for every wordnet, the distribution of nominal *Lsynsets* vs. *Isynsets*, grouped by synset

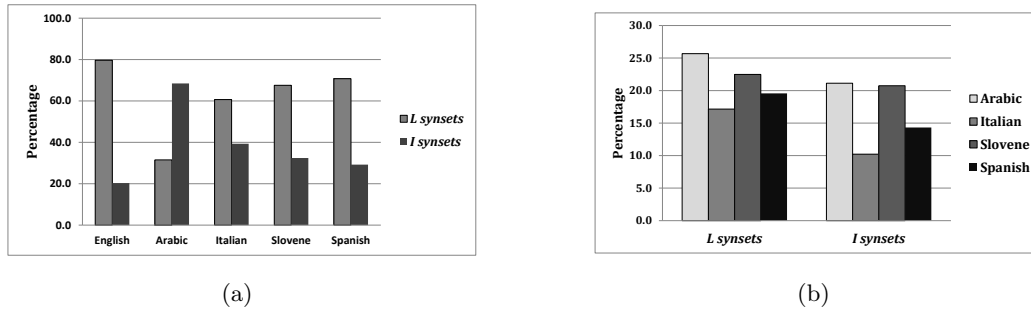


Figure 8: Percentage of Leaf synsets (*Lsynsets*) and Intermediate synsets (*Isynsets*) (a) in the gold standards (b) in the non-covered synsets

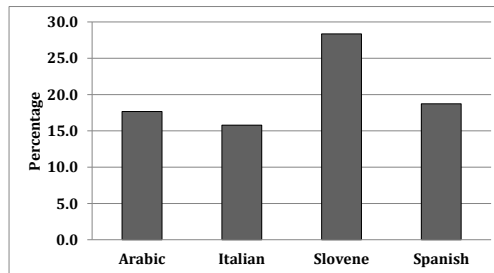


Figure 9: Neighbour synset coverage for non-covered synset

category. We can notice that *Lsynsets* are more likely to be not covered than *Isynsets*, and that a large number of non-covered synsets consists of synonymless synsets.

Moreover, we would like to evaluate if, for non covered synsets, translations return candidate matches that are at least *semantically similar* to their equivalent synsets in the target language. Neighbor synsets (i.e., hypernyms, hyponyms, and siblings) are usually considered similar in many wordnet and graph based similarity measures (Navigli, 2009). Inspired by work presented by Resnik and Yarowsky (1999), one could consider establishing a weighted mapping between a synset in a source language and synsets in the target language, such that the weight represents the degree of similarity between the source and the target synset. In our experiment we define similar synsets as one being either hyponym/hypernym or siblings of the other one. As shown in Figure 9, the average percentage of synsets not covered with *MT* for which at least one synset similar to the equivalent synset is found among the candidate matches is 20.1%. This is consistent with the intuition that machine translation systems provide translations which (implicitly) capture a more coarse-grained sense specification than the fine-grained sense specification encoded in the wordnets. In fact, it was observed that WordNet is sometimes too fine-grained even for human judges to agree (Hovy, Marcus, Palmer, Ramshaw, & Weischede, 2006).

Observation 14. For a significant percentage of non covered synsets (20.1%, on average), machine translation tools return synsets that are at least similar to their equivalent synsets in the target language.

Table 14: Distribution of leaf and intermediate (non-)covered synsets by category

Synsets	Arabic				Italian			
	Non-covered		Covered		Non-Covered		Covered	
	Leaf	Inter	Leaf	Inter	Leaf	Inter	Leaf	Inter
M-MWS	9.1	16.0	35.0	39.9	5.3	1.5	71.6	21.5
M-OWS	14.3	27.1	27.0	31.6	11.4	4.0	60.3	24.3
MIX	3.7	8.5	19.6	68.2	4.1	2.1	44.6	49.3
P-MWS	2.7	5.2	17.9	74.2	4.0	2.6	34.4	59.0
P-OWS	8.8	13.1	19.7	58.4	11.3	5.5	44.1	39.0
<i>ALL</i>	7.8	14.4	23.5	54.3	8.6	3.6	52.1	35.7

Synsets	Slovene				Spanish			
	Non-covered		Covered		Non-Covered		Covered	
	Leaf	Inter	Leaf	Inter	Leaf	Inter	Leaf	Inter
M-MWS	10.9	4.7	60.0	24.3	8.0	1.1	82.5	8.4
M-OWS	13.2	4.6	70.4	11.9	17.8	4.2	64.9	13.1
MIX	11.3	7.1	38.1	43.5	9.6	2.4	48.7	39.3
P-MWS	13.0	8.3	28.1	50.6	7.1	3.3	28.3	61.3
P-OWS	22.3	10.2	35.3	32.2	14.1	7.0	37.0	41.8
<i>ALL</i>	15.2	6.7	52.4	25.7	13.8	4.2	56.9	25.0

Based on this observation, the candidate match retrieval step can be modified so as to include among the candidate matches also synsets similar to the ones retrieved by translation. This approach has been followed by several cross-lingual ontology matching systems (Fu et al., 2012; Cruz, Palmonari, Caimi, & Stroe, 2013; Faria et al., 2014). However, expanding the set of considered candidate matches has the disadvantage of increasing the difficulty of mapping selection task. The results of our analyses suggest that the expansion of the candidate matches set is a technique that could be applied only to particular categories of source synsets, e.g., to *synonymless leaf* synsets. This could provide a system (or a user, in an interactive matching settings) with a greater ability to map synsets that are less likely to be covered by translations, without increasing the number of candidate matches for every source synset, e.g., for synsets that have *distinguishable monosemous* candidate matches (see *Observation 8*).

6.4 Lessons Learned & Future Works

In this section we summarize the main results and findings of our study and highlight some potential future directions.

A general conclusion that we draw from our study is that machine translation tools and multilingual knowledge resources return useful translations for a very large number of concepts. Thus, translations provide a valuable support for candidate match retrieval in cross-lingual ontology matching, covering from a minimum of 75.2% to a maximum of 89.9% synsets in the four languages other than English considered in this study. If we consider that BabelNet also incorporates translations derived from mappings in the Open Multilingual Wordnet (Bond & Foster, 2013) (which have been excluded in our study because they have been used as gold standards), this coverage is expected to even increase for several resource-rich languages covered by this wordnet. In addition, our experiments suggest that translations can be helpful, to a more limited extent, and for selected categories of synsets, also in the mapping selection task.

Concisely, the main results of our experiments suggest that:

- monosemous concepts (i.e., concepts that have only monosemous words) are considered to be more domain-specific;
- combining lexical resources improves the quality of results;
- machine translation tools perform poorer on domain-specific concepts than on domain-independent ones;
- synonymful synsets have higher coverage than synonymless synsets;
- most, but not all, monosemous concepts can be mapped confidently even with simple selection methods (e.g., translation-based majority voting);
- mappings involving polysemous but synonymless synsets are harder to filter out within the mapping selection task;
- the more the coverage for synonym words (in synonymful synsets), the easier is the mapping selection task.

Compared to previous systems, which used machine translation tools considering only one translation direction, in our study we built dictionaries that cover both translation directions by including reverse translations. This technique has been shown to significantly improve the coverage of translations. In practice, candidate matches can be found for a larger number of input concepts, thus increasing the upper-bound recall for cross-lingual ontology matching systems. As a promising future research direction, one may try to further improve coverage by considering additional information available in machine translation tools like Google Translate (e.g., reverse translation synonym-like sets, part-of-speech tagged translations, and translation scores). Such additional information can increase not only the upper-bound recall, but also the precision, if adequately used in the matching selection step. For example, one may compare the words returned by reverse translations with an input source synset, e.g., by using our translation-correctness measure (Eq.5). The translation with higher translation-correctness could be given a higher weight in the selection step.

The selection of a correct mapping from a set of candidate matches still remains a difficult task, in particular when *contextual knowledge* cannot be used to disambiguate the meaning of the concepts. However, the findings of our paper suggest several research directions that can mitigate this problem.

On one hand, the simple baseline selection method based on majority voting used in our experiments should be overcome by more sophisticated methods. For example, in a recent work, we define a lexical similarity measure based on evidence collected from translations and we run a local similarity optimization algorithm to improve the assignments between source and target concepts (Abu Helou & Palmonari, 2015). In future work, we would like to leverage the analysis of mapping selection difficulty as dependent on the lexical characterisation of source and target concepts (e.g., polysemous vs. monosemous concepts, or synonymless vs. synonymful synsets) discussed in this paper. We plan to investigate matching algorithms that could adapt their behavior based on the category of the source synset and its candidate matches.

On the other hand, some cross-lingual mappings may still be hard to decide upon using a fully automatic approach. Thus, we would like to investigate in the cross-lingual ontology

matching domain, the adoption of semi-automatic matching methods. A web application could be used to solve difficult cross-lingual matching tasks, as the one proposed to match short service descriptions in different languages (Narducci, Palmonari, & Semeraro, 2013). Beyond this, interactive matching processes that aggregate inputs given by a multiplicity of users, either experts (Cruz et al., 2014) or crowd workers (Sarasua et al., 2012) seem particularly promising in large cross-lingual matching tasks. The findings of this paper are particularly useful for similar approaches because they can help to decide on which mappings the user inputs are more valuable (e.g., polysemous and synonymless concepts). Overall we plan to follow the latter research directions to use a map model to ease the construction of a “lexical-semantic” ontology in the context of the Arabic Ontology Project (Abu Helou et al., 2014), which also motivated the study presented in this paper.

7. Conclusions

In this study we have investigated the effectiveness of automatic translations derived from a state-of-the-art machine translation tool (Google Translate) and a state-of-the-art multilingual knowledge resource (BabelNet) to support cross-lingual ontology mapping. To perform our analysis we used four very large repositories of cross-lingual mappings, which include mappings from wordnets in four different languages to the English WordNet. Effectiveness of automatic translations is analyzed in terms of coverage and correctness. One key contribution of our study, besides the scale of the experiments, is the analysis of the effectiveness of automatic translations for specific categories of synsets.

For example, we found that automatic translations achieve lower coverage for domain specific concepts. As another example, we found that the amount of monosemous words that are correctly translated into polysemous words in another language is not negligible: cross-lingual ontology mapping methods that use the monosemous word heuristic may lead to include a several wrong mappings in an alignment. At a coarse grain, our analyses suggest that automatic translations are capable of covering a large number of word senses, in particular when more multilingual lexical resources (e.g., Google Translate and BabelNet) and translation strategies (i.e., the reverse translations of Google Translate) are integrated. On the other hand, automatic translations are correct only to a limited extent, at least when compared to translations derived from manually mapped wordnets.

The analyses discussed in this paper inspired the definition of a cross-lingual similarity measure for lexical ontologies (Abu Helou & Palmonari, 2015). A natural subsequent step is to further utilize the study outcomes in cross-lingual mapping systems. One promising research direction is to define adaptive mapping methods such that different strategies are used depending on the lexical characterization of the source concepts. For example, one could integrate interactive mapping methods or crowdsourcing approaches to decide about a subset of the mappings, which are estimated to be particularly difficult to map. Another research direction that we plan to investigate is a method to estimate of concept ambiguity in small ontologies that do not explicitly contain synonyms, e.g., by matching them against wordnets. Such a method would help us to use adaptive cross-lingual mapping methods on axiomatic ontologies or other lexically-poor data sources, e.g., web tables.

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments and valuable suggestions, and Pikakshi Manchanda for her help in proofreading. This work was supported in part by COMSODE project (FP7-ICT-611358) and SIERA project (FP7-INCO-295006). Corresponding author: Mamoun Abu Helou, E-mail: mamoun.abuhelou@disco.unimib.it.

References

- Abu Helou, M. (2014). Towards constructing linguistic ontologies: Mapping framework and preliminary experimental analysis. In *Proceedings of the Second Doctoral Workshop in Artificial Intelligence, Pisa, Italy*. CEUR-WS.
- Abu Helou, M., & Palmonari, M. (2015). Cross-lingual lexical matching with word translation and local similarity optimization. In *Proceedings of the 10th International Conference on Semantic Systems, SEMANTiCS 2015, Vienna, Austria, September*.
- Abu Helou, M., Palmonari, M., Jarrar, M., & Fellbaum, C. (2014). Towards building linguistic ontology via cross-language matching. In *Proceedings of the 7th International Conference on Global WordNet*.
- AGROVOC (2014). Multilingual agricultural thesaurus. <http://aims.fao.org/vest-registry/vocabularies/agrovoc>.
- Al-Kabi, M. N., Hailat, T. M., Al-Shawakfa, E. M., & Alsmadi, I. M. (2013). Evaluating English to Arabic Machine Translation Using BLEU. *International Journal of Advanced Computer Science and Applications(IJACSA)*, 4(1).
- Apertium (2015). Open-source machine translation platform. <http://www.apertium.org>.
- Apidianaki, M. (2009). Data-driven semantic analysis for multilingual wsd and lexical selection in translation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (ACL)*, EACL '09, pp. 77–85, Stroudsburg, PA, USA. Association for Computational Linguistics (ACL).
- BabelNet (2012). Very large multilingual semantic network. <http://babelnet.org>.
- Bentivogli, L., Forner, P., Magnini, B., & Pianta, E. (2004). Revising the wordnet domains hierarchy: Semantics, coverage and balancing. In *Proceedings of the Workshop on Multilingual Linguistic Ressources*, MLR '04, pp. 101–108, Stroudsburg, PA, USA. Association for Computational Linguistics (ACL).
- Bing (2016). Bing translate. <http://www.bing.com/translator>.
- Birzeit (2011). Arabic Ontology. <http://sina.birzeit.edu/ArabicOntology/>.
- Bond, F., & Foster, R. (2013). Linking and extending an open multilingual wordnet. In *ACL (1)*, pp. 1352–1362. The Association for Computer Linguistics (ACL).
- Bouma, G. (2010). Cross-lingual ontology alignment using eurowordnet and wikipedia. In *LREC*. European Language Resources Association (ACL).

- Boyd-Graber, J., Osherson, & Schapire, R. (2006). Adding dense, weighted connections to WordNet connections to wordnet. In *Proceedings of the Third Global WordNet Meeting*.
- Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, 25(6), 657–674.
- CAT (2014). Chinese Agricultural Thesaurus. <http://www.ciard.net/partners/labof-chinese-agricultural-ontology-services>.
- COMSODE (2015). Components supporting the open data exploitation. <http://www.comsode.eu/>.
- Costa-jussà, M. R., Farrús, M., Marino, J. B., & Fonollosa, J. A. R. (2012). Study and comparison of rule-based and statistical catalan-spanish machine translation systems. *Computing and informatics*, 31(2), 245–270.
- Cruz, I. F., Palmonari, M., Caimi, F., & Stroe, C. (2013). Building linked ontologies with high precision using subclass mapping discovery. *Artif. Intell. Rev.*, 40(2), 127–145.
- Cruz, I., Loprete, F., Palmonari, M., Stroe, C., & Taheri, A. (2014). Pay-as-you-go multi-user feedback model for ontology matching. In *Knowledge Engineering and Knowledge Management*, Vol. 8876 of *Lecture Notes in Computer Science*, pp. 80–96. Springer.
- De Melo, G., & Weikum, G. (2009). Towards a universal wordnet by learning from combined evidence. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, Hong Kong, China, November*, pp. 513–522. ACM.
- De Melo, G., & Weikum, G. (2012). Constructing and utilizing wordnets using statistical methods. *Language Resources and Evaluation*, 46(2), 287–311.
- Djeddi, W. E., & Khadir, M. T. (2014). Xmap++: Results for oaei 2014. In *Proceedings of the 9th International Workshop on Ontology Matching co-located with the 13th International Semantic Web Conference (ISWC 2014), October*, Vol. 20, pp. 163–169.
- Els, L., & Véronique, H. (2010). Semeval-2010 task 3: Cross-lingual word sense disambiguation. In *Proceedings of the 5th International Workshop on Semantic Evaluation*. Association for Computational Linguistics (ACL).
- EUROVOC (2015). EU’s multilingual thesaurus. <http://europa.eu/eurovoc>.
- Faria, D., Martins, C., Nanavaty, A., Taheri, A., Pesquita, C., Santos, E., Cruz, I. F., & Couto, F. M. (2014). Agreementmakerlight results for oaei 2014. In *Proceedings of the 9th International Workshop on Ontology Matching co-located with the 13th International Semantic Web Conference (ISWC 2014), October*, Vol. 20, pp. 126–134.
- Fellbaum, C. (Ed.). (1998). *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA; London.
- Fišer, D. (2007). Leveraging parallel corpora and existing wordnets for automatic construction of the slovene wordnet. In *LTC*, Vol. 5603 of *Lecture Notes in Computer Science*, pp. 359–368. Springer.
- Fodor, J. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

- Fu, B., Brennan, R., & O’Sullivan, D. (2009). Cross-lingual ontology mapping - an investigation of the impact of machine translation. In *ASWC*, pp. 1–15.
- Fu, B., Brennan, R., & O’Sullivan, D. (2012). A configurable translation-based cross-lingual ontology mapping system to adjust mapping outcomes. *J. Web Semantic*, 15, 15–36.
- Gale, W. A., Church, K. W., & Yarowsky, D. (1992). Using bilingual materials to develop word sense disambiguation methods. In *In Proceedings of the International Conference on Theoretical and Methodological Issues in Machine Translation*.
- Gonzalez-Agirre, A., Laparra, E., & Rigau, G. (2012). Multilingual central repository version 3.0. In *LREC*, pp. 2525–2529. European Language Resources Association (ELRA).
- Google (2015). Google Translate. <https://translate.google.com/>.
- Gracia, J., Montiel-Ponsoda, E., Cimiano, P., Gómez-Pérez, A., Buitelaar, P., & McCrae, J. (2012). Challenges for the multilingual web of data. *Web Semantic*, 11, 63–71.
- Hertling, S., & Paulheim, H. (2012). Wikimatch - using wikipedia for ontology matching. In *Ontology Matching*, Vol. 946 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Hirst, G. (2004). Ontology and the lexicon. In *Handbook on Ontologies and Information Systems*. Springer.
- Horrocks, I. (2008). Ontologies and the semantic web. *Communications of the ACM*, 51(12), 58–67.
- Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., & Weischede, R. (2006). Ontonotes: the 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL. Companion Volume, Short Papers on XX, NAACL 06. Association for Computational Linguistics (ACL)*, Morristown, NJ, USA.
- Hovy, E., Navigli, R., & Ponzetto, S. P. (2012). Collaboratively built semi-structured content and artificial intelligence: The story so far. *Artificial Intelligence*.
- Ide, N., Erjavec, T., & Tufis, D. (2002). Sense discrimination with parallel corpora. In *Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions - Volume 8, WSD ’02*, pp. 61–66, Stroudsburg, PA, USA. Association for Computational Linguistics (ACL).
- Jarrar, M. (2006). Position paper: Towards the notion of gloss, and the adoption of linguistic resources in formal ontology engineering. In *Proceedings of the 15th International Conference on World Wide Web*, pp. 497–503, NY, USA. ACM.
- Jarrar, M. (2011). Building a formal arabic ontology (invited paper). In *Proceedings of the Experts Meeting On Arabic Ontologies And Semantic Networks. Alecso, Arab League. 26-28 July. Tunis*.
- Jarrar, M., Yahya, A., Salhi, A., Abu Helou, M., Sayrafi, B., Arar, M., Daher, J., Hicks, A., Fellbaum, C., Bortoli, S., Bouquet, P., Costa, R., Roche, C., & Palmonari, M. (2014). Arabization and multilingual knowledge sharing- final report on research setup. In *SIERA Project 2.3 Deliverable*.

- Jiménez-Ruiz, E., Grau, B. C., Xia, W., Solimando, A., Chen, X., Cross, V., Gong, Y., Zhang, S., & Chennai-Thiagarajan, A. (2014). Logmap family results for oaei 2014?. In *Proceedings of the 9th International Workshop on Ontology Matching co-located with the 13th International Semantic Web Conference (ISWC 2014), October*, Vol. 20, pp. 126–134.
- Jiménez-Ruiz, E., & Grau, B. C. (2011). Logmap: Logic-based and scalable ontology matching. In *The Semantic Web–ISWC 2011*, pp. 273–288. Springer.
- Json-W3schools (2015). Javascript object notation. <http://www.w3schools.com/json/>.
- Lefever, E., & Hoste, V. (2013). Semeval-2013 task 10: Cross-lingual word sense disambiguation. *Proc. of SemEval*, 158–166.
- Liang, A. C., & Sini, M. (2006). Mapping agrovoc and the chinese agricultural thesaurus: Definitions, tools, procedures. *The New Review of Hypermedia and Multimedia*, 12(1), 51–62.
- Lin, F., Butters, J., Sandkuhl, K., & Ciravegna, F. (2010). Context-based ontology matching: Concept and application cases. In *10th IEEE International Conference on Computer and Information Technology, CIT 2010, Bradford, West Yorkshire, UK, June 29-July 1, 2010*, pp. 1292–1298.
- Lin, F., & Krizhanovsky, A. (2011). Multilingual ontology matching based on wiktionary data accessible via sparql endpoint. *CoRR*, abs/1109.0732.
- LOGD (2015). Linking Open Government Data. <http://logd.tw.rpi.edu/iogds-data-analytics>. [Online; accessed March-2015].
- Margolis, E., & Laurence, S. (2014). Concepts. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring edition).
- Mariò, J. B., Banchs, R. E., Crego, J. M., de Gispert, A., Lambert, P., Fonollosa, J. A. R., & Costa-jussà, M. R. (2006). N-gram-based machine translation. *Comput. Linguist.*, 32(4), 527–549.
- McCarley, J. S. (1999). Should we translate the documents or the queries in cross-language information retrieval?. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics (ACL)*, ACL '99, pp. 208–214, Stroudsburg, PA, USA. Association for Computational Linguistics (ACL).
- McCarthy, D., & Navigli, R. (2009). The english lexical substitution task. *Language Resources and Evaluation*, 43(2), 139–159.
- MCR (2012). Multilingual Central Repository. <http://adimen.si.ehu.es/web/MCR>.
- Meilicke, C., Garca-Castro, R., Freitas, F., van Hage, W. R., Montiel-Ponsoda, E., de Azevedo, R. R., Stuckenschmidt, H., Svab-Zamazal, O., Svatek, V., Tamin, A., Trojahn, C., & Wang, S. (2012). Multifarm: A benchmark for multilingual ontology matching. *Web Semantics: Science, Services and Agents on the World Wide Web*, 15(3).
- Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM*, 38(11), 39–41.

- Miller, G. A., Leacock, C., Teng, R., & Bunker, R. T. (1993). A semantic concordance. In *Proceedings of the workshop on Human Language Technology, HLT '93*, pp. 303–308, Stroudsburg, PA, USA. Association for Computational Linguistics (ACL).
- Mulwad, V., Finin, T., & Joshi, A. (2013). Semantic message passing for generating linked data from tables. In *International Semantic Web Conference (1)*, Vol. 8218 of *Lecture Notes in Computer Science*, pp. 363–378. Springer.
- Narducci, F., Palmonari, M., & Semeraro, G. (2013). Cross-language semantic matching for discovering links to e-gov services in the lod cloud. In *KNOW@LOD*, Vol. 992 of *CEUR Workshop Proceedings*, pp. 21–32. CEUR-WS.org.
- Nastase, V., Strube, M., Boerschinger, B., Zirn, C., & Elghafari, A. (2010). Wikinet: A very large scale multi-lingual concept network. In *LREC*. European Language Resources Association (ACL).
- Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Comput. Surv.*, 41(2), 10:1–10:69.
- Navigli, R., & Ponzetto, S. P. (2012). Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193(0), 217 – 250.
- OAEI (2015). The ontology alignment evaluation initiative. <http://oei.ontologymatching.org>.
- Oliver, A., & Climent, S. (2012). Building wordnets by machine translation of sense tagged corpora. In *Proceedings of the 6th International Conference on Global WordNet*.
- Omegawiki (2015). Wikimedia. <http://www.omegawiki.org>.
- OMWN (2015). The open multilingual wordnet. <http://compling.hss.ntu.edu.sg/omw/>.
- Otero-Cerdeira, L., Rodriguez-Martinez, F. J., & Gmez-Rodriguez, A. (2015). Ontology matching: A literature review. *Expert Systems with Applications*, 42(2), 949 – 971.
- OWL (2004). Web ontology language. <http://www.w3.org/TR/owl-features/>.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL)*, ACL '02, pp. 311–318, Stroudsburg, PA, USA. Association for Computational Linguistics (ACL).
- Pianta, E., Bentivogli, L., & Girardi, C. (2002). Multiwordnet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*.
- Pinker, S. (1994). *The Language Instinct*. Harper Perennial Modern Classics, New York.
- Po, L., & Sorrentino, S. (2011). Automatic generation of probabilistic relationships for improving schema matching. *Information Systems*, 36(2), 192 – 208. Special Issue: Semantic Integration of Data, Multimedia, and Services.
- RDF (2014). Resource Description Framework. <http://www.w3.org/RDF/>.
- RDFS (2014). RDF Schema. <http://www.w3.org/TR/rdf-schema>.

- Resnik, P., & Yarowsky, D. (1999). Distinguishing systems and distinguishing senses: new evaluation methods for word sense disambiguation. *Natural Language Engineering*, 5(2), 113–133.
- Rodríguez, H., Farwell, D., Farreres, J., Bertran, M., Martí, M. A., Black, W., Elkateb, S., Kirk, J., Vossen, P., & Fellbaum, C. (2008). Arabic wordnet: Current state and future extensions. In *Proceedings of the Forth International Conference on Global WordNet*.
- Sag, I., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword expressions: A pain in the neck for nlp. In *Computational Linguistics and Intelligent Text Processing*, Vol. 2276 of *Lecture Notes in Computer Science*, pp. 1–15. Springer.
- Sagot, B., & Fišer, D. (2008). Building a free french wordnet from multilingual resources. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC)*, Marrakech, Maroc.
- Sarasua, C., Simperl, E., & Noy, N. (2012). Crowdmap: Crowdsourcing ontology alignment with microtasks. In *The Semantic Web ISWC 2012*, Vol. 7649 of *Lecture Notes in Computer Science*, pp. 525–541. Springer.
- Saveski, M., & Trajkovski, I. (2010). Automatic construction of wordnets by using machine translation and language modeling. In *In Proceedings of Seventh Language Technologies Conference, 13th International Multiconference Information Society, volume C*.
- Shvaiko, P., & Euzénat, J. (2013). Ontology matching: State of the art and future challenges. *IEEE Trans. Knowl. Data Eng.*, 25(1), 158–176.
- Shvaiko, P., Euzénat, J., Mao, M., Jiménez-Ruiz, E., Li, J., & Ngonga, A. (Eds.). (2014). *Proceedings of the 9th International Workshop on Ontology Matching collocated with the 13th International Semantic Web Conference (ISWC 2014)*, Riva del Garda, Trentino, Italy, October 20, 2014, Vol. 1317 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Sorrentino, S., Bergamaschi, S., Gawinecki, M., & Po, L. (2010). Schema label normalization for improving schema matching. *Data & Knowledge Engineering*, 69(12), 1254 – 1273. Special issue on 28th International Conference on Conceptual Modeling (ER 2009).
- Spelke, S. E. (2003). What makes us smart? core knowledge and natural language. In *Language in Mind: Advances in the Study of Language and Thought*, pp. 277–311. Mit Press.
- Spohr, D., Hollink, L., & Cimiano, P. (2011). A machine learning approach to multilingual and cross-lingual ontology matching. In *Proceedings of the 10th International Conference on The Semantic Web - Volume Part I, ISWC’11*, pp. 665–680. Springer.
- Suchanek, F. M., Kasneci, G., & Weikum, G. (2008). Yago: A large ontology from wikipedia and wordnet. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(3), 203–217.
- Tomaž, E., & Fišer, D. (2006). Building slovene wordnet. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC’06)*, Genoa, Italy.

- Trojahn, C., Fu, B., Zamazal, O., & Ritze, D. (2014). State-of-the-art in multilingual and cross-lingual ontology matching. In *Towards the Multilingual Semantic Web*, pp. 119–135. Springer.
- Tufis, D., Cristea, D., & Stamou, S. (2004). Balkanet: Aims, methods, results and perspectives. a general overview. *Special Issue on BalkaNet. Romanian Journal on Science and Technology of Information*.
- Vincent, V. A. (2013). Macro- and micro-averaged evaluation measures [[basic draft]]. In *Technical report*.
- Vossen, P. (2004). Eurowordnet: A multilingual database of autonomous and language-specific wordnets connected via an inter-lingualindex. *International Journal of Lexicography*, 17(2), 161–173.
- Vossen, P., Eneko, A., Francis, B., Wauter, B., Axel, H., Amanda, H., Shu-Kai, H., Hitoshi, I., Chu-Ren, H., Kyoko, K., Andrea, M., German, R., Francesco, R., Roxane, S., & Maurizio, T. (2010). *KYOTO: A Wiki for Establishing Semantic Interoperability for Knowledge Sharing Across Languages and Cultures*, pp. 265–294. Springer.
- Wikidata (2015). Wikimedia. <http://www.wikidata.org>.
- Wikipedia (2015a). List of languages by number of native speakers. <http://en.wikipedia.org/wiki/List-of-languages-by-number-of-native-speakers>.
- Wikipedia (2015b). Wikipedia. <https://www.wikipedia.org>.
- Wikispecies (2015). Wikimedia. <https://species.wikimedia.org>.
- Wiktionary (2015). Wikimedia. <https://www.wiktionary.org>.
- WordNet-Princeton (2015). Semantically tagged glosses. <http://wordnet.princeton.edu/glosstag.shtml>.
- Zhang, Z. (2014). Towards efficient and effective semantic table interpretation. In *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I*, pp. 487–502.
- Zhuge, H., Xing, Y., & Shi, P. (2008). Resource space model, owl and database: Mapping and integration. *ACM Transactions on Internet Technology (TOIT)*, 8(4), 20.